

# KNOWLEDGE, REPRESENTATION, AND THE PHYSICAL WORLD

by

**Michael N. Tamir**

B.A. Mathematics, Columbia University, 2003

B.A. Philosophy, Columbia University, 2003

M.S. Physics, University of Pittsburgh, 2007

M.A. Mathematics, University of Pittsburgh, 2008

Submitted to the Graduate Faculty of  
the Department of Philosophy in partial fulfillment  
of the requirements for the degree of

**Doctor of Philosophy**

University of Pittsburgh

2012

UNIVERSITY OF PITTSBURGH  
DEPARTMENT OF PHILOSOPHY

This dissertation was presented

by

Michael N. Tamir

It was defended on

August 3rd 2012

and approved by

John D. Norton, PhD, History & Philosophy of Science

John Earman, PhD, History & Philosophy of Science

James Woodward, PhD, History & Philosophy of Science

Robert Batterman, PhD, Philosophy

Gordon Belot, PhD, Philosophy (University of Michigan)

Dissertation Advisors: John D. Norton, PhD, History & Philosophy of Science,

John Earman, PhD, History & Philosophy of Science

# Knowledge, Representation, and the Physical World

Copyright © by Michael N. Tamir

2012

# Knowledge, Representation, and the Physical World

Michael N. Tamir, PhD

University of Pittsburgh, 2012

## Abstract

This dissertation answers how mathematical representations enable knowledge of physical systems. Contemporary responses rely on matching the properties of physical systems to properties in mathematical models, arguing that such matching allows scientists to successfully draw conclusions about physical systems through the inspection of their models. We argue that such “matching accounts” cannot adapt to the routine *mismatching* pervasive in physical theories. These mismatching problems arise both when idealized models match some “similar” but better behaved potential physical system, and in cases we classify as *pathological* idealization, where the models employed must satisfy constraints that *could not possibly be matched by realistic physical systems* (e.g. requiring an infinite particle number or infinite density). In the latter cases such pathological constraints can also lead to incompatibilities with the governing laws of the physical theory. Despite such pathologies, conclusions drawn with these representations seem to enable improved understanding and empirically confirmable knowledge of the studied physical systems.

To address this dichotomy, we develop a novel condition of successful mathematical representation, called  $\epsilon$ -fidelity, under which mismatched models may facilitate knowledge of realistic physical systems. Arguing against direct matching, we propose that representations can meet the conditions of  $\epsilon$ -fidelity by establishing a manifold of associations between topological *neighborhoods* of mathematical models and *clusters* of relevantly similar physical systems. We then demonstrate that this shift in the scope of representation relationships explains how suitably similar models entail conclusions about the relevant systems while avoiding the problems of individual model to system mismatching.

As a signature case study, we investigate Einstein’s canonical interpretation of the *geodesic principle*, originally proposed to govern how gravitating bodies travel according to general relativity theory. We argue that under the canonical interpretation models of bodies must either meet unrealistic assumptions or violate the theory’s fundamental field equations, marking them as pathological idealizations. To recover the principle, we reinterpret geodesic dynamics as a *universality thesis* about the collective behavior of certain classes of systems, explaining how this reinterpretation satisfies the  $\epsilon$ -fidelity criteria and can be used to gain knowledge about the observable motion of actual classes of gravitating bodies.

## Acknowledgments

I would like to thank each of the members of my committee for their assistance and continual support in researching the current project. I would especially like to thank my committee co-chairs John Norton and John Earman for their support and advice throughout this process. The influence of each committee member's own work on the following has been significant. I am grateful both for this influence and for the many conversations with each member on how my project intersects with their work. I am also indebted to Balázs Gyenis, particularly with respect to the material of [chapter 2](#). Much of my thinking on this subject has benefited greatly from our numerous conversations on the topic of mathematical representation.

# Contents

<b>1</b>	<b>Mathematics, Representation, and Scientific Inference</b>	<b>1</b>
1.1	Faithful Representation and the Goals of Science . . . . .	3
1.1.1	What is Representation? . . . . .	3
1.1.2	Inferential Surrogates and Faithful Representation . . . . .	10
1.2	Defending Fidelity . . . . .	17
1.2.1	The Mysterious Fidelity Problem . . . . .	17
1.2.2	Matching, Maps and Morphisms . . . . .	19
1.2.3	The Bridge “Structure” of Königsberg . . . . .	25
1.3	Ampliative Abstraction and Epistemic Debt . . . . .	33
1.3.1	The Observation of Data vs. Phenomena . . . . .	34
1.3.2	Determining the Data: Rejecting Naivety . . . . .	39
1.3.3	Determining the Data: Temperature . . . . .	42
1.3.4	Data to Phenomena Reasoning . . . . .	49
<b>2</b>	<b>Scientific Representation and <math>\epsilon</math>-Fidelity</b>	<b>57</b>
2.1	Idealization and Plato’s Problem . . . . .	59
2.1.1	Framing the Problem . . . . .	59

# CONTENTS

2.1.2	Scope and Structure . . . . .	64
2.1.3	Close Enough for Carpentry . . . . .	71
2.2	Solving Plato’s Problem . . . . .	75
2.2.1	$\epsilon$ -Faithful Representation . . . . .	75
2.2.2	$\epsilon$ -Fidelity and Experimentation . . . . .	82
2.2.3	Putting the $\epsilon$ in Subtle Transition . . . . .	85
2.3	$\epsilon$ -Fidelity and Proximity to Pathology . . . . .	90
2.3.1	Limits and Anchor Properties . . . . .	90
2.3.2	Closeness in Context . . . . .	96
2.4	Concluding Remarks: Towards the Point of Pathologies . . . . .	103
<b>3</b>	<b>Proving the Principle</b>	<b>107</b>
3.1	Einstein and The Canonical Account . . . . .	109
3.1.1	Geodesic Dynamics . . . . .	109
3.1.2	Whither Test Particles? . . . . .	114
3.1.3	Einstein and Grommer’s “Three Ways” to get it Straight . . . . .	117
3.2	Singularity Proofs . . . . .	124
3.2.1	The Geodesic that Wasn’t There . . . . .	124
3.2.2	Distributional Energy-momentum . . . . .	128
3.3	$0^{th}$ -Order Proofs . . . . .	135
3.4	Limit Operation Proofs . . . . .	141
3.4.1	Geroch-Jang Particles . . . . .	141
3.4.2	Ehlers-Geroch Particles . . . . .	145
3.5	Towards a Geodesic Universality Thesis . . . . .	149



# CONTENTS

<b>4</b>	<b>Geodicity and <math>\epsilon</math>-Faithful Universality</b>	<b>151</b>
4.1	Another place for the Principle? . . . . .	153
4.2	Universality in Physics . . . . .	157
4.2.1	The Paradigm Case: Universality in Phase Transitions . . . . .	158
4.2.2	The Same but Different: Analyzing Universality . . . . .	162
4.3	The Geodesic Universality Thesis . . . . .	166
4.3.1	The Similarity and Diversity of Geodesic Universality . . . . .	166
4.3.2	Explaining Geodesic Universality . . . . .	168
4.4	Explanation without Reification . . . . .	171
<b>5</b>	<b>Conclusion</b>	<b>175</b>
	<b>Appendix A Tensor Distributions</b>	<b>178</b>
	<b>Appendix B Proof from Generalized Conservation of <math>T^{ab}</math></b>	<b>181</b>
	<b>Appendix C GT-regular and Semi-regular Metrics</b>	<b>185</b>
	<b>Bibliography</b>	<b>188</b>

# List of Figures

1.1.1 <i>Visual representation of Obama, 2008</i> . . . . .	8
1.2.1 <i>Adapted Map of Königsberg, 1613.</i> . . . .	26
1.2.2 <i>Representation of Königsberg Bridge Structure.</i> . . . .	26
1.3.1 <i>Thermodynamic <math>P</math>-<math>V</math> curve</i> . . . . .	53
1.3.2 <i>Possible <math>P</math>-<math>V</math> data records</i> . . . . .	53
2.1.1 <i>Physical wooden frame.</i> . . . .	59
2.1.2 <i>Quadrilaterals <math>\mathbf{A}</math> and <math>\mathbf{B}</math> with congruent side lengths.</i> . . . .	60
2.1.3 <i>“C-shaped” inference schema</i> . . . . .	62
2.1.4 <i>Diagram of the Pincock <math>\varphi_4</math> mapping</i> . . . . .	68
2.1.5 <i>Diagram of the “wide scope” <math>\Phi</math> mapping</i> . . . . .	69
2.1.6 <i>Member of the set <math>\mathcal{V}_{\delta,S}</math></i> . . . . .	74
2.2.1 <i>Diagram of <math>\delta</math> to <math>\epsilon</math> deduction</i> . . . . .	78
2.3.1 <i>Solution to Plato’s original challenge</i> . . . . .	92

LIST OF FIGURES

2.3.2 <i>The rationallium challenge</i> . . . . .	92
2.3.3 <i>Graph of magnetization density curves <math>m_N(h)</math></i> . . . . .	99
4.2.1 <i>Phase diagram of a generic material at fixed density.</i> . . . .	159
4.2.2 <i>Adapted plot of (Guggenheim, 1945) data rescaled for criticality.</i>	160

# Chapter 1

## Mathematics, Representation, and Scientific Inference

*At this point an enigma presents itself, which in all ages has agitated inquiring minds. How can it be that mathematics, being after all a product of human thought which is independent of experience, is so admirably appropriate to the objects of reality?*

- Albert Einstein ([Einstein, 1922a](#))

The seduction of applying mathematics to the physical world is compelling: A carpenter wishes to lay some baseboard along a three meter long hallway, but she only has four meter long baseboard stock. The stock will not fit. To solve this dilemma, she removes a meter from the end of the baseboard stock and proceeds to lay the now three meter long baseboard. How did she know this would work?

It seems undeniable that mathematics was involved in the carpenter's solution. We want to say that she “applied arithmetic” to solve her problem. This sort of instance is so terribly prosaic we take the fact that such applications should work completely for granted. If the

success of such instances are so easily acceptable, we should be able to answer why it works so well. How did the carpenter know what to do?

Without reflection upon what mathematical claims are supposed to be about, the application of these claims appears deceptively unproblematic. Philosophical complications can arise with the innocent observation that mathematics is abstract. Mathematical inferences are made about abstract systems of relations on a domain of abstract mathematical objects. *Mathematics is not supposed to be about the physical properties of concrete material objects.* Arithmetic is about numbers and their interrelations, not about wooden baseboards. So how do such completely abstract inferences like ‘ $4 - x = 3$  only if  $x = 1$ ’ work so well when laying oversized baseboard stock? This question deserves an answer not just in prosaic cases of carpentry, but in the often far more complex and exotic applications of mathematics in scientific theories. For sophisticated applications of mathematics in physical theories, the question becomes vital to understanding if we can trust such applications at all. This question is precisely Albert Einstein’s “enigma” referred to in the opening quote. How do inferences about abstract mathematical systems allow us to make inferences about the physical world? How can math apply?

In the first half of this work (chapters 1 and 2), we will develop an account of successful mathematical representation germane, in particular, to mathematical applications used to gain scientific knowledge of our physical world. In chapters 3 and 4, we will then immediately proceed to put our account to work uncovering and then resolving a potential pathology in the dynamics of Einstein’s general theory of relativity. We shall argue that the example found in Einstein’s theory is paradigmatic of many complex and potentially problematic mathematical representations actually advanced in well accepted physical theories. We hence present the analysis of these latter chapters as a framework for resolving these complications

## 1.1. FAITHFUL REPRESENTATION AND THE GOALS OF SCIENCE

and recovering epistemologically fruitful analyses of such scientific representations.

In this chapter we begin to develop our account of successful mathematical representation in science as follows: In section 1.1 we will review the current literature on the theory of representation. We will also argue (section 1.1.2) that the driving criteria of *scientific* representation should depend on the knowledge gained about the targets of scientific study. In section 1.2 we identify the first of two major problems facing any account of mathematical representation aimed at meeting these epistemic goals, called the *mysterious fidelity problem*. We proceed by reviewing and then critically analyzing potential options available for resolving this challenge often discussed in the literature. In section 1.3, using the *data-phenomena* distinction of Bogen & Woodward (1988), we argue that the legitimate targets of mathematical representations in science should be scientific *phenomena*. We then conclude with a review of some of the epistemological challenges that result from attempting to identify well defined abstractions of such targets. It will ultimately be shown that many of the epistemological challenges raised in this chapter can be eliminated or at least mitigated with our ultimate account of successful mathematical representation in science, completed in chapter 2.

## 1.1 Faithful Representation and the Goals of Science

### 1.1.1 What is Representation?

The term ‘representation’ has suffered from much confusion. This confusion has manifested in what Sorin Bangu (2009, note 5) recently described as an unfortunately “bewildering variety of uses in recent work in the philosophy of science.” This absence of uniformity in the literature is a symptom of the fact that there is likewise a significant lack of a univocal usage

### 1.1. FAITHFUL REPRESENTATION AND THE GOALS OF SCIENCE

of the term in natural language. The term ‘representation’ has multiple legitimate usages. Nonetheless, (if adequately sanitized) it should be evident that something aptly referred to as “representation” is going on when we use mathematical models or equations to communicate or think about certain features of physical systems. It is hence worth considering what representation is (and is not), and if there exists a subspecies of the concept at work when using mathematics in physical sciences.

To begin, every representation must have something that is supposed to be represented, often referred to as the *target* of the representation, and something doing the representing, often referred to as the *vehicle*. So in the case of a tourist map depicting the footpaths of Prague, the map itself plays the role of the vehicle of representation and its target is the paths located in that part of that city in eastern Europe. Of course, even this feature of representation has its hiccups. By saying that a representation *has* a target, we must not require that the target *exists* either now or at any other time. For example, a painting of Pegasus can still legitimately represent the mythical winged horse even though the creature does not and has never existed.

Though vehicles of representations may very well have features in common with their targets, *a representation is not reducible to a mere similarity or resemblance relation holding between any subset of properties or relations true of both the vehicle and its target*. For one thing, *misrepresentations* are representations too. For example, imagine we attend a political rally and come across a poster depiction of some United States president dressed as Hitler and that the poster is claiming that “president *X* is Hitler,” such a depiction still may legitimately be taken as a representation of the president despite the strong manifest dissimilarity between any United States president and the leader of the Third Reich. Or consider the presidential seal, which shares dramatically few properties with any United

### 1.1. FAITHFUL REPRESENTATION AND THE GOALS OF SCIENCE

States president. Even when the vehicle fails to resemble or directly misleads as in these cases, representation is possible. We may conclude from the overwhelming abundance of examples such as these that a condition of sufficiently significant resemblance it is *not necessary* for every form of successful representation.

Similarity also *fails to be sufficient* for representation. For instance, I am quite similar (identical in fact) to myself, yet (save perhaps in an instance of *pro se* legal defense) it would be inappropriate to ever claim that “I represent myself.” Or consider the power sander in my basement shop: It is a quite popular model sander made by a reliable company capable of manufacturing power tools with negligible variation in their construction. It is safe to say that there exist many other sanders in the world that share virtually every property of material construction with the sander in my shop, yet it would be a mistake to think that these facts alone suffice for a representation relation to obtain between my sander and one of its sander brethren.

Nelson Goodman (1976, pp3-10) classically argued that representation also lacks the “logical properties” that are characteristic of resemblance (or similarity). As already illustrated by some of the above examples, representation is not generally symmetric or reflexive whereas resemblance relations are always symmetric and reflexive. Everything always resembles itself and if two things resemble each other there is no directionality to the resemblance. On the other hand, except in rare cases (e.g. certain paintings by René Magritte), it would be a mistake to say that representation relations generally hold in which the vehicle and target are identical. And almost universally it is a mistake to infer that if  $X$  represents  $Y$  then  $Y$  likewise represents  $X$ . So, for example, though a painting of the Eiffel tower may represent the Eiffel tower, we are hard pressed to say that the Eiffel tower also represents the painting. There is an inherent directionality to representation, distinguishing the vehicle from the



target.

Comparison with resemblance or similarity relations highlights the need for a key component that has come to be taken (at least implicitly) as necessary for any representation relation.<sup>1</sup> Unlike resemblance or similarity relations, which might be said to obtain “in a vacuum” a representation relation obtaining between a vehicle and its target can happen only by the design of a *representing agent* or *agents*. In each of the examples of successful representation just mentioned, the implicit existence of a representing agent available to use or treat the vehicle as representative plays a necessary role for a representation relation to obtain. The tourist map of Prague only represents Prague if its creator designed it as such. If Putnam’s famous ant incidentally were to trace out lines geometrically similar to footpaths of Prague with mapping-ink dipped ant legs on mapping paper, the resulting product would not be a representation. Putnam’s ant cannot provide the requisite agency. Hence, the resulting sheet, even if it is ink-drop for ink-drop identical to my tourist map, would fail to be a representation of the footpaths in the eastern European city. The Hitler-president poster held by the protester is representative because we surmise that the protester created the image for the purpose of representing a president as being Hitler-like and the imagery encourages viewers to take the poster as being so representative (whether they agree with the content or not). In the case of my favorite sander, if I were to display it to my neighbor, proudly suggesting that it is a shining example of the quality associated with Acme Tool orbit sanders, we might then, in this context (unlike before), say that my usage of the tool as such enables us to identify it as now representative of its Acme Tool brethren.

Bas van Fraassen has recently focused on this component of representation as the paramount element required for a representation to obtain, recommending the following (somewhat cir-

---

<sup>1</sup>See e.g. (van Fraassen, 2008, Contessa, 2007, Suárez, 2003, Giere, 2004).

## 1.1. FAITHFUL REPRESENTATION AND THE GOALS OF SCIENCE

cular) necessary condition for successfully representation:<sup>2</sup>

There is no representation except in the sense that some things are used, made or taken, to represent something as thus or so. (van Fraassen, 2008, p23)

This condition suggests that purposeful action of an agent or agents to represent something is a necessary condition for any representation to occur.<sup>3</sup> Following this observation then, representation is (at least) a three place relation, in which an agent or agents  $A$  use a vehicle  $V$  to represent a target  $T$ . The agency required in a representation can, of course, occasionally be taken as implicit despite its significance in this relation. A reason for this might be that the representer providing the agency of the representation is the one representing, or that the community of agents for whom the vehicle in question is taken as representative is obvious from the context. Even if only implicit, however, agency plays an essential role in any representation relation, the presence of which informs some of the important distinctions mentioned above (such as directedness) between representation and mere resemblance or similarity.

A final (and also often implicit) component of the representation relation we will refer to as the *content* of a representation. The content of a representation is what the target is being *represented as* by the vehicle. Not every property of a vehicle of representation directly communicates the content of the representation intended by a representing agent. For instance, there might be incidental material properties of a representing vehicle that

---

<sup>2</sup>Though he is careful to explain that he does not promote any “theory of representation,” van Fraassen nonetheless suggests that “if he did” the quoted sentence would be his “*Hauptsatz*.”

<sup>3</sup>One ambiguity worth noting but not dwelling on concerns where the agency has to come from. Does the agent need to be the creator of the vehicle? In many cases this does seem to be a prime source of relevant agency in a representation. However, in cases of multi-agent representation created and used by many individuals (as perhaps in scientific representations), it is not clear that the initial creator’s intentions are paramount. Moreover, in cases where the creator’s original intentions are inaccessible (for example, of works of art created long ago where it is unclear who the original artist even was) it still seems fair to suggest that a representation occurs if an audience member is available to provide the agency of the representation in, say, interpreting the ostensible vehicle’s target. In general, any user of the vehicle as representative of a target can count as providing the requisite agency.



Figure 1.1.1: *Shepard Fairey's representation of Obama for the 2008 presidential election.*

have nothing to do with its representation of the target. It would be foolish, for example, to think that our tourist map representation of the footpaths of Prague suggests that the actual paths are smooth (like paper), or that they may be folded up and placed in our pockets. It is not just that these claims are untrue about the actual footpaths of Prague (in that case we would simply have a *misrepresentation*). Rather, they are not even part of the content of the representation in question. Not only can mere incidental material properties of a vehicle of representation be irrelevant to the content, but properties of the vehicle that actually play a role in indirectly communicating the representational content can nonetheless fail to be directly part of the content of the representation. For instance, consider the depiction of Barack Obama in the well known Shepard Fairey poster displayed in figure 1.1.1. It would be an error to infer that part of the content of the poster's representation of Obama as a candidate includes the claim that his face is red, white, and blue. Again, it is not just that such a claim is manifestly false (Obama's face has never been these unnatural colors, and definitely not during his 2008 candidacy for president). The claim plays no part whatsoever in the properties of Obama that the Fairey poster is meant to represent. Note the subtlety:

### 1.1. FAITHFUL REPRESENTATION AND THE GOALS OF SCIENCE

the representational content of the poster presumably has to do with characterizing Obama as a patriotic candidate, and this point is in part achieved by the very selection of colors of the American flag in question that are used in Fairey's depiction of Obama. However, an interpretation of the image depicting the candidate Obama as *literally* red, white, and blue is not a part of the representational content.

There is no evident algorithm for identifying the precise representational content of an arbitrary representation. Instead the precise representational content of a given representation appears to be highly case sensitive and must be evaluated on a specific representation by specific representation basis. It might be observed that the representing agent or agents and their intended interpretation or their usage play a central role in adjudicating the appropriate content of a representation. However, this is not to say that in any act of representation, the content of the representation must be articulable in explicit discursive form even by the representing agent (or agents). This fact is again exemplified by the Obama poster or, say, by Pablo Picasso's *Guernica* (or any number of other of works of art). In each of these works it is not necessarily comprehensively clear what the precise explicit representational content of the the respective works is (though we might have some ideas). As they say, a picture is worth a lot of words. Nonetheless, it is clear that however the content might be explicitly articulated (if at all), it must be distinguished from the vehicle of representation itself.

At this point we have identified a minimum of four components of any representation relation:

(AVTC) Agent  $A$  uses vehicle  $V$  to represent target  $T$  as  $C$ .

where  $C$  is the content of the representation. In every identification of a representation relation all four components of the (AVTC) need not be explicitly referred to in contexts

## 1.1. FAITHFUL REPRESENTATION AND THE GOALS OF SCIENCE

where one or more of the roles is obvious or, in the case of the content, not (entirely) articulable.<sup>4</sup> Nevertheless, all four components (perhaps more)<sup>5</sup> at minimum play distinct and important roles in any representation relation.

### 1.1.2 Inferential Surrogates and Faithful Representation

We have now distinguished between representation and resemblance and have forestalled inclinations to treat representation as strictly reducible to the intrinsic properties and relations held by the target and the vehicle alone. Representation can occur with almost no specific resemblance between the vehicle and its target whatsoever. Imagine for instance that while going for a pensive walk on the beach an agent, Alice, pauses, picks up an unremarkable stone and turning to her walking companion solemnly states that “this stone represents the totality of all my fears and concerns about the future,” and subsequently hurls the stone into the sea in a cathartic and metaphorical release of those worries. In this story the stone may count as representative of her concerns (at least for Alice). Representation can be used for such catharsis if one likes, and in these cases there appears to be absolutely no reason why the vehicle (like Alice’s worry stone) must share anything in common with the target (all her worries). Other examples of denotational representation established by fiat or convention

---

<sup>4</sup>Of course, when the content of a representation has some particular emotive component that is non-discursive, it will be impossible to explicitly capture in the *C* slot of **(AVTC)** such emotive content. Since non-discursive elements of certain (e.g. artistic) representations nonetheless constitute potentially legitimate representational content, it would be a mistake to eliminate their possibility despite the complications associated in capturing such content discursively in the **(AVTC)** form as given. Note, however, the primary point of identifying the **(AVTC)** form is to highlight that there are (at least) four components to a representation. And, the observation that part of the *C* component may or may not be discursive does not in any way suggest that this point is illegitimate even in such complicating cases.

<sup>5</sup>Instead of representational content, [Giere \(2004, p743\)](#) identifies the agent’s *purpose* as a key fourth component of representation. It is arguable that the purpose of the agent in representing the target with a representation’s vehicle can be subsumed as part of the content of the representation. That being said, there is no problem in allowing for the possibility of a fifth (or for that matter sixth, or seventh, etc.) component of representational relations should there be recalcitrant examples in which this is not possible. For our current purposes, it is enough to identify that there are *at least* the four components identified by the **(AVTC)** format. See also ([van Fraassen, 2008, p21](#)).

### 1.1. FAITHFUL REPRESENTATION AND THE GOALS OF SCIENCE

abound. The presidential seal, the Canadian flag, and countless other symbols in certain contexts and with certain audiences serve as vehicles of representation which share few if any properties with their targets. Nonetheless, there are plenty of alternative genuine examples of representation beyond such cases of conventional or fiat denotation.

Representations can be used, for example, to communicate thoughts or induce emotions, as in some of the artistic examples already mentioned. In these cases, it appears that the precise properties of the vehicle may well have relevance to the purpose of the representation. Further, representation instances such as the Prague map example might have specific practical purposes, namely, figuring out how to get around Prague. Such practical instances of representation appear especially germane to the case of scientific representation because the representational vehicle is used in specific ways to improve our understanding of the target. Hence, we have reason to look at this particular species of usage in investigating the general kind of representation employed in science.

One prominent family of theories in the philosophy of science literature on representation can be grouped together as what might be called *inferential accounts of representation*.<sup>6</sup> Each account varies in its details, but the key feature of inferential accounts lies in their attention to the use of a representation vehicle as a kind of *surrogate* for inferences we wish to draw about target itself.<sup>7</sup> In his inferential account of representation, Mauricio Suárez (2004, p773) proposes the following condition on representation:

**(Sur)**  $V$  represents  $T$  only if  $V$  allows competent and informed agents to draw specific inferences regarding  $T$ .

---

<sup>6</sup>See e.g. (Swoyer, 1991, Hughes, 1997, Suárez, 2003, Contessa, 2007, Bueno & Colyvan, 2011).

<sup>7</sup>The term *surrogative reasoning* was first introduced by Chris Swoyer (1991) and has subsequently been adopted by a number of inferential accounts of representation. Swoyer's original usage of the term was somewhat less generalized (focusing specifically on cases of what he calls "structural representation") than the current usage in the literature.

### 1.1. FAITHFUL REPRESENTATION AND THE GOALS OF SCIENCE

In cases of representation where **(Sur)** is satisfied, the vehicle of representation acts as an *inferential surrogate* for the target of the representation, and the process of drawing such inferences is referred to as *surrogative reasoning*. A paradigmatic case of inferential surrogacy occurs with our Prague map representation. Inferences about where to make particular turns, or where certain landmarks are located with respect to a location can be made in the case of the Prague map representation. These inferences are not made by direct inspection of or reflection about the target itself (Prague), but by inspection of the various features of the representational vehicle (the paper map). In this way, the vehicle indeed acts as a kind of surrogate for inferences we wish to make about the actual target (hence the terminology).

As already exemplified by the examples of conventional denotation mentioned above, an inferential account of representation insisting that the *only* legitimate instances of representation that ever occur are those in which inferential surrogacy is possible are bound to fail. (E.g. considering the presidential seal will allow for vanishingly few inferences about the United States presidency.) However, it is clear that the ability to enable surrogative reasoning is rather vital to cases of *scientific* representation. Scientific study in whatever form it takes aims to increase our knowledge and understanding of the phenomena studied in that discipline. Scientists seek to form judgments (ideally “correct” ones in one respect or another) about the phenomena studied. Though the particular details about what kind of judgments can or should be formed (e.g. if we can infer facts about unobservables from experimental data or not), what the phenomena being studied is (unobservable entities, experimental data in particular scenarios, etc.), and what the standards or modes of correctness might be (e.g. literal truth, empirical adequacy, adequacy within certain experimental error), the following thesis should hold uncontroversially:

### 1.1. FAITHFUL REPRESENTATION AND THE GOALS OF SCIENCE

- (G) The primary goal of scientific investigation is to form judgments that increase our knowledge or understanding of the kind of phenomena studied.

If a scientific representation is intended in any way to contribute to this primary goal, then it is immediate that enabling us to form judgments about the phenomena studied is required of any useful scientific representation. In other words, any scientific representation potentially useful to the primary goal of science must act as an inferential surrogate for the phenomena represented. It must allow us to draw conclusions and form judgments about the studied phenomena.<sup>8</sup> Hence, though inferential surrogacy is demonstrably not necessary for representations of all types, *any scientific representation that can be put to use for the primary goal of science must enable surrogate reasoning.*

Gabriele Contessa (2007) has recently offered admirable clarification on the subject of scientific representation by drawing some significant terminological distinctions between certain varieties of representation often conflated in the literature. Using the concept of inferential surrogacy, he distinguishes three categories of representation. The weakest being referred to as *denotational representation*. Paradigm cases of (mere) denotational representations include the examples mentioned at the opening of this subsection such as the worry stone, the presidential seal, or the Canadian flag. Denotational representations still take the (at least) four component form of (AVTC), but particularly in cases where the representation is merely denotational, there may be vanishingly little representational content (e.g. though the presidential seal represents something like the United States presidency, there is little representational content beyond the signification itself). Merely denotational representations are hence highly dependent on conventional or fiat signification of the target by the vehicle

---

<sup>8</sup>To be clear, it is consistent with the thesis suggested here that such conclusions be restricted in various ways should one adopt an especially restrictive outlook on the possible domain of scientific knowledge (e.g. we can only know about “observables”). The thesis is hence entirely neutral on the epistemic issues of sufficient justification under contention in typical scientific realism debates.



### 1.1. FAITHFUL REPRESENTATION AND THE GOALS OF SCIENCE

established by an agent or community of agents employing the representation. The next variety Contessa identifies is *epistemic representation*. Epistemic representations are denotational in that their vehicles signify their targets, but they also enable surrogate reasoning about their targets. Contessa characterizes epistemic representation as follows:

A vehicle is an *epistemic representation* of a certain target for a certain user if and only if the user is able to perform valid (though not necessarily sound) surrogate inferences from the vehicle to the target. (Contessa, 2007, p53)

In order to understand what an epistemic representation is, we must hence take a look at what it takes to make a surrogate inference valid.<sup>9</sup> The general requirements for the validity of a surrogate inference seem to be suggested in Suárez’s (**Sur**) where he requires that the specific surrogate inferences must be made by “competent and informed agents” using the representation. The requirement of competent and informed agency should imply that the representing agents must have an understanding of the content of the representation being used. We might then say that valid surrogate inferences are those (and only those) drawn about the target of a representation by an agent who has an understanding of which features of the vehicle are (and are not) relevant to the representation. In other words, valid surrogate inferences are the kind of inferences drawn by agents using the representational content “in the way it is supposed to be used.” Hence, the ability to draw valid surrogate inferences is directly dependent on the content of a representation.

We already (indirectly) observed the important distinction between valid and invalid surrogate reasoning while discussing the need for representational content in the first place. In discussing the Prague map representation or the Obama poster, we then noted that not every feature of the vehicle of representation is relevant to how the target is being repre-

---

<sup>9</sup>Contessa does not give an independent definition of validity in the case of inferential surrogacy, but instead builds the notion into certain rules for the specific kind of “interpretational” account of epistemic representation that he later develops (Contessa, 2007, p61).

### 1.1. FAITHFUL REPRESENTATION AND THE GOALS OF SCIENCE

sented. Taking advantage of our current terminology we can now say that inferring from the Prague map that the footpaths of Prague are smooth like paper, or inferring from the poster that Obama’s face is red, white, and blue count as *invalid* surrogate inferences. *To draw such inferences would involve substantive confusion over what counts as the representational content of the respective representations.* On the other hand, inferring from the map that walking a certain sequence of paths enables one to get to a particular location in Prague, or inferring from the poster that (according to Fairey’s interpretation of patriotism) it would be patriotic to vote for Obama in the 2008 presidential election, are examples of *valid* surrogate inferences. Drawing these conclusions about the respective targets in these cases involves appropriate understanding and usage of their representational contents.

Determining if a valid surrogate inference has been made is quite sensitive to the particular representation in question. Just as it would be a mistake to think that there exists a universal algorithm for determining the content of an arbitrary representation, it is a mistake to think that there is a universal test of validity for any given surrogate inference independent of the specific content of the representation in question. In such cases substantive work done by competent users to clarify the content of representations is required to adjudicate between the valid and invalid inferences.<sup>10</sup> That being said, despite the case sensitivity of validity evaluation, there remains a key point to observe: *An understanding of representational content is essential to successful determination of valid inferential surrogacy.*

The final variety of representation that Contessa identifies consists of those epistemic representations which not only enable inferential surrogacy, but enable *surrogate inferences whose conclusions are correct about the represented target.* This category, called *faithful epistemic representation* (or *faithful representation* for short) is meant to designate the usages

---

<sup>10</sup>In large part, this is the sort of project embarked on in chapters 3 and 4.

### 1.1. FAITHFUL REPRESENTATION AND THE GOALS OF SCIENCE

where the term ‘represents’ is intended as a kind of “success verb” (Contessa, 2007, p54). The terminology distinguishes between epistemic representations that end up *misrepresenting* their targets (e.g. a depiction of some US president as Hitler) and those that are successful (e.g. an up to date Prague map). Recall, in the above analysis of epistemic representation, Contessa carefully notes that though such representations need to enable valid inferential surrogacy, such inferences need not be sound. The differences between valid and sound inferential surrogacy lies in whether or not the inferred conclusion is true. *Sound surrogate inferences* are valid surrogate inferences with conclusions that are true of their target (Contessa, 2007, p51). Faithful epistemic representations are those epistemic representations with sound surrogate inferences.

Of course, epistemic representations will typically enable more than one valid surrogate inference. In general, an epistemic representation might enable some sound surrogate inferences and some unsound (but valid) surrogate inferences. Hence, Contessa explains that “unlike epistemic representation, faithful epistemic representation is a matter of degree,” the idea being that whereas any enabling of surrogate reasoning qualifies a representation as being epistemic, some epistemic representations are more faithful than others (Contessa, 2007, p55).<sup>11</sup> Extreme cases of faithful epistemic representation in which every valid inference is also sound he refers to as *completely faithful epistemic representations*, and the other extreme where every valid inference is unsound he dubs *completely unfaithful epistemic representations*. All other epistemic representations are cases of *partial faithfulness* (and partial misrepresentation).

---

<sup>11</sup>Of course, the idea that it is a matter of “degree” suggests the existence of a total ordering of faithful epistemic representations (presumably) indicating “how many” or “what proportion” of valid surrogate inferences are sound. Though the possibility of such an ordering is suspect, nonetheless, the general distinction he points to is legitimate.

## 1.2 Defending Fidelity

### 1.2.1 The Mysterious Fidelity Problem

With Contessa's distinctions in hand, we can now concisely state what is desired in scientific representations of physical phenomena. *We want our scientific representations to be as epistemically faithful as possible with respect to the features of the physical systems studied by scientists.* Of course, there may be significant disagreement (amongst philosophers of science at least) over what the proper features of the physical systems studied by scientists indeed are: Whereas an anti-realist may suggest that these features include only that which is experimentally observable, a realist might suggest that unobservable features (e.g. entity existence, some kind of general structuralist features, etc.) are also properly included. What is important is that we want our representations to be epistemically faithful to the studied features (whatever they happen to be). In particular, when physicists use mathematics, we would like the mathematical representation to be as epistemically faithful to the studied features of the physical phenomena as possible. These criteria should appear eminently reasonable. When scientists represent the physical world with math we want their representations to actually tell us something about the world and we want the things they tell us to be correct.

Two significant problems face these desiderata for mathematical representations of the physical world. The first problem brings us back to what we referred to as Einstein's enigma in the introduction to this chapter. Contessa remarks on the difficulty that faces an exclusively inferential account of representation as follows:

On the inferential conception, the user's ability to perform inferences from a vehicle to a target seems to be a brute fact, which has no deeper explanation. This makes the connection between epistemic representation and valid surrogate reasoning needlessly

## 1.2. DEFENDING FIDELITY

obscure and the performance of valid surrogate inferences an activity as mysterious and unfathomable as soothsaying or divination. (Contessa, 2007, p61)

If a scientific representation is representative purely in virtue of the epistemic faithfulness that it engenders, then the soundness of the inferences drawn using the vehicle about the target is entirely mysterious. If we happen to have a faithful representation, perhaps we know that the inferences made by means of the vehicle work, but we have no understanding of why the representation does work or more importantly, why we can continue to expect it to work in novel situations.

This problem is especially treacherous in the case of mathematical representations of physical phenomena. Unless one wants to reject a view of mathematics according to which mathematical claims are not *a priori* but instead established empirically through some sort of interaction with the concrete physical world,<sup>12</sup> one must concede that mathematics is not specifically about any targeted phenomena. But if the mathematics is inherently not about any physical phenomena (particularly the phenomena targeted), then when a mathematical representation of a physical phenomena happens to be epistemically faithful, such success is genuinely mystifying in the absence of further justification. Arithmetic is not about baseboard stock, but how then does it function so well for the carpenter making surrogate inferences about how much to cut? Let us call this problem facing any faithful mathematical representation of the physical world the *problem of mysterious fidelity*. In this section we will look at how some accounts of representation, motivated by the mysterious fidelity problem (if not explicitly then at least by its looming stench), attempt to offer such a justification.<sup>13</sup>

---

<sup>12</sup>The canonical example of this sort of radical position was of course proposed by John Stuart Mill (1986).

<sup>13</sup>Accounts from Bueno & Colyvan (2011) or Hughes (1997) explicitly try to incorporate some kind of mapping or denotation into their respective inferential schemas, presumably because of a concern over this problem. Other accounts such as (Pincock, 2004) apparently take the need for granted and begin by directly

## 1.2. DEFENDING FIDELITY

The second problem, resolved in chapter 2, we will refer to as *Plato’s problem*. This problem will result from certain attempts to solve the first problem. We will argue that standard attempts to address the mysterious fidelity problem are uniformly foiled by recurring failures resulting from the fact that mathematical concepts are, in a sense, “too precise” to be well matched to the phenomena studied by physicists. As we shall see, all attempts offered so far in the literature to solve the mysterious fidelity problem fail to account for the pervasive use of idealization (broadly construed) in physicists’ mathematical representations of physical phenomena.

### 1.2.2 Matching, Maps and Morphisms

A potential response to the mysterious fidelity problem rests on the idea that though mathematical representations use purely abstract mathematics for surrogate inference, in some way the properties and relations of the (non-abstract) targeted physical system are captured and simulated by the purely mathematical structures. Supposedly, when there is a certain kind of similarity between the conditions placed on the abstract mathematical system<sup>14</sup> and the “actual” properties and relations of the targeted physical system, inferences drawn from such conditions in the case of the vehicle should hold in the case of the target.

Of course, as already exhibited by the numerous non-epistemic and non-faithful representations developing what might serve as a solution. That something like the fidelity mystery facing a purely inferential account (as in (Suárez, 2004)) motivates the development of a mapping or denotational picture (even when an inferential surrogacy has been incorporated) is broadly active in these cases. See also Swoyer’s original structuralist account of surrogate reasoning in (Swoyer, 1991).

<sup>14</sup>Here and in what follows, by *mathematical system* we will mean a domain (set) of abstract elements with a set of relations defined on elements in that domain. The *structure* of a mathematical system refers precisely to the set of relations defined on the domain of that system. So, for instance, the integers modulo 3 (i.e.  $(\{0, 1, 2\}; +)$ ) constitutes a mathematical system, whose structure is the congruence relations of addition *mod* 3 (e.g.  $1 + 2 \equiv 0$ ). As we shall investigate below, this crisply defined concept of structure in the case of a mathematical system is not to be conflated with the rather amorphous notion of “physical structure” or “structure of a physical system,” which has suffered unfortunate obfuscation in philosophical literature (particularly the scientific realism literature).

## 1.2. DEFENDING FIDELITY

tations discussed in section 1.1.1, relations like similarity or resemblance are neither necessary nor sufficient for representation. However, according to the response in question, when it comes to *faithful* epistemic representations, some sort of similarity condition or conditions may potentially eliminate the otherwise looming threat of mysterious fidelity. So, how does such a fidelity through similarity argument look in detail and what kind of similarity is supposed to be sufficient?

To understand how the similarity defense gains traction, let us consider examples of faithful representations where the vehicle is not mathematical. The potential justification of fidelity through similarity exists, for instance, in cases such as accurate (geographical) maps. Why does our Prague map work? An answer no doubt can be given by the observation that the way symbols indicating pathways and landmarks are oriented on the paper map somehow “matches” the way that actual physical pathways and landmarks of Prague happen to be related to one another: The map was designed so that pathways intersect on the map if and only if the pathways in Prague to which they refer intersect, landmarks are signified on the map adjacent to a pathway symbol if and only if the physical landmark is located on the referred to path. And, assuming the map is to scale, we can even say that the ratio of the lengths of two pathways on the map is the same as the ratio of the distances along the two referenced pathways in the city. If these sorts of resemblances (among others) hold, between the vehicle map and the target city, then we want to say that “inferences drawn (exclusively) from the properties and relations *shared* by both the target and the vehicle should hold true not only for the vehicle, but also for the target.”

But what is meant by “shared” properties and relations? Certainly not that the properties are literally the same of the map and the city. Imagine a marked pathway signifying street X on a map intersects another marked pathway signifying street Y, and that moreover, the

## 1.2. DEFENDING FIDELITY

actual signified street X intersects the signified street Y in the actual city. The markings on the map signifying street X and the concrete street X do not actually possess the same property or relation. Unlike the concrete street X, the markings on the map signifying street X do not really intersect the concrete street Y, instead they intersect a line on the map. Similarly, unlike the markings on the map signifying street X, the concrete street X does not intersect the line on the map referring to street Y, instead it intersects the concrete street Y. What is shared is not the very same properties or relations, but a kind of *analogy* of relations: The relation between the signified streets X and Y is shared by the map and the city because the symbols on the map signifying X and Y intersect in an analogous way. It is the presence of such analogous relations between the vehicle and the target that allow inferences about the target (such as “if I want to get to path X, I can continue on path Y”) to be drawn by inspecting the analogous relations depicted by the vehicle of representation, the paper map.

To take another example, imagine that there is a clock face pictured on our Prague map in the region of the map labeled ‘old town square,’ we can then say that the presence of the clock face in that region mimics the city even though it is false that there is literally a image of a clock pictured on the cobble stones of the actual city’s old town. Instead, what is located in that region of the city is a physical landmark (the old town hall) which is supposed to be signified by the clock picture. What is “shared” by the target and vehicle is not the precise properties or relations of either the target or the vehicle in question but the analogous relations between the respective parts of the target and vehicle in question. So, the old town hall is located in the old town square in the physical city, and analogously, the clock picture is located in the region of the map labeled ‘old town square.’ It is the analogy of the relation between the spatial localization of the symbols in the vehicle (the map) to



## 1.2. DEFENDING FIDELITY

the relation of their referents in the target (the city) that the target and vehicle share. The availability of this sort of analogy between interrelations of the target and the vehicle allows us to inspect the map in order to draw the (surrogate) inference that if we travel to the old town square we will be able to visit the old town hall.

Observing that such an “analogy of relations” potentially plays a pivotal role in figuring out why faithful epistemic relations work, many analyses of representation in the literature have endeavored to precisify this concept. The most prominent such family of attempts are referred to as *mapping accounts* of mathematical representation.<sup>15</sup> A paradigm example of a mapping account has been recently developed by Chris Pincock, who characterizes mathematical representation as follows:<sup>16</sup>

[A] wholly mathematical model represents a physical situation in virtue of a structure-preserving mapping like an isomorphism or an homomorphism between the physical situation and the mathematical model. (Pincock, 2007b, p960)

The term ‘mapping’ in this context is suggestive of the mathematical concept of a *mapping function* which assigns a unique output for each input. In the exclusively mathematical context, referring to a mapping as establishing a homomorphism or an isomorphism is well defined. A mapping from a structured set  $S$  onto a structured set  $T$  is homomorphic, if the relations among the elements of  $S$  are “preserved” when we look at their image in the set  $T$  (two elements are related in the structure of  $S$  only if they are mapped to elements that are also so related in the structure of  $T$ ). To take a simple example, consider the structured sets  $S := (\{even, odd\}; +)$  and the structured set  $T := (\{positive, negative\}; \times)$ . We know from

---

<sup>15</sup>Recent mapping accounts include (Bueno & Colyvan, 2011, Pincock, 2004, 2007a,b, Hughes, 1997), and more classic considerations of mapping (or in the case of Reichenbach proto-mapping) accounts include (Putnam, 1978, Lewis, 1984, Reichenbach, 1965).

<sup>16</sup>Pincock like many mapping theorists does not explicitly distinguish between faithful, epistemic, and non-epistemic representation. However, it might be inferred (at least implicitly) from the context of his discussions of representations, which primarily focus on scientific application and explanation, that the variety of representation he is analyzing is not representation in general but the faithful variety in particular.

## 1.2. DEFENDING FIDELITY

grade school that the rules for adding even and odd numbers are related by four 3-place “addition” relations in the following way:

$$\begin{aligned} \text{even} &= \text{even} + \text{even} \\ \text{even} &= \text{odd} + \text{odd} \\ \text{odd} &= \text{odd} + \text{even} \\ \text{odd} &= \text{even} + \text{odd} \end{aligned} \tag{1.2.1}$$

Likewise we know that positive and negative numbers are related by four 3-place “multiplication” relations in the following way:

$$\begin{aligned} \text{positive} &= \text{positive} \times \text{positive} \\ \text{positive} &= \text{negative} \times \text{negative} \\ \text{negative} &= \text{negative} \times \text{positive} \\ \text{negative} &= \text{positive} \times \text{negative} \end{aligned} \tag{1.2.2}$$

A quick inspection of these sets of relations (i.e. their respective structures) reveals that if we define a mapping  $\varphi$  from the elements of  $\mathbf{S}$  to the elements of  $\mathbf{T}$  as follows:

$$\varphi \begin{cases} even & \mapsto positive \\ odd & \mapsto negative \end{cases}$$

then the elements in the set  $S$  are related by the addition relations (1.2.1) only if the elements in  $T$  to which they are mapped are likewise related by a corresponding multiplication relation (1.2.2). It is in this sense that we can say that the structure (1.2.1) on  $S$  is *preserved* under the mapping by the multiplication relations (1.2.2) in the structure of  $T$ .<sup>17</sup>

It is not hard to see the attraction of appealing to the mathematical notion of homomorphic mapping in order to bring some precision to an account of the kind of “analogy of relations” discussed above in the faithful (geographical) map example. The sort of relations preservation that we get from a homomorphism between two structured mathematical sets seems to be the sort of “analogy of relations” of the respective mathematical structures that we wanted. However, it has recently been argued by van Fraassen that (despite such ostensible potential) a mapping between a mathematical system and the physical world involves a kind of category error:

If the target is not a mathematical object, then we do not have a well-defined range for the function, so how can we speak of an embedding or isomorphism or homomorphism

---

<sup>17</sup>The mapping  $\varphi$  establishes an isomorphism between our two structured sets because not only is the structure of  $S$  preserved in  $T$  under the mapping, but also there is a one to one correspondence (bijection) between the respective elements (each element of  $T$  has a unique partner in  $S$  that gets sent to it by  $\varphi$ ). This one to one correspondence is not generally required for homomorphisms. A simple example of this consists of the mapping from the positive integers with addition to our structured set  $S$  in which every even number is sent to *even* and every odd number is sent to *odd*. If we take two numbers that will be sent to *even* (e.g. 6 and 28), then their sum (34) will also be a number sent to *even* in accordance with the first relation of (1.2.1). Similarly other combinations of arbitrary even or odd positive integers will be in accordance with one of the four relations of (1.2.1). Hence the addition relations over the positive integers are preserved in the structure (1.2.1) of  $S$ , making the mapping homomorphic. However, this mapping from positive integers to  $S$  is not one to one, because multiple positive integers get sent to the respective elements *even* and *odd*. So such a mapping is an example of an homomorphism that is not an isomorphism.

## 1.2. DEFENDING FIDELITY

or whatever between the target and some mathematical object? (van Fraassen, 2008, p241)<sup>18</sup>

The potential of such an appeal to the concept of homomorphic mappings unravels under the observation that unlike strictly mathematical contexts, in the abstract structured set (e.g.  $S$ ) to another (e.g.  $T$ ) but instead attempt to establish a mapping between an abstract mathematical structure (i.e. the theoretical model(s)) and a concrete non-mathematical target, part of the physical world. In contrast to an abstractly defined mathematical system, determined by strictly defined sets of conditions and relations on the elements of the system, the physical world does not come predefined as a structured domain. This is not to say that there are not certainly all sorts of relations that might be legitimately attributed to parts of physical systems or groups of systems, but unlike mathematical systems, the relations of the parts of the physical systems differ fundamentally.

### 1.2.3 The Bridge “Structure” of Königsberg

Mapping accounts appear to succumb to a category error in assuming that physical systems can be treated like mathematical systems, but in a sense such a point may strike us as “merely” a technicality. We can imagine a mapping account proponent replying “so what if the relations of a physical systems are not actual *mathematical relations*, if they are still true of (parts of) the system, then shouldn’t we be able to preserve such physical structure (i.e. the relations) in a mathematical representation by means of a homomorphism-like correspondence of parts of the physical system to parts of the mathematical system?” In

---

<sup>18</sup>It should be noted that the “direction” of the mapping is not all ways uniform in the literature. Some mapping accounts (e.g. (Bueno & Colyvan, 2011, Hughes, 1997)) portray the mapping as going from the physical world to the model, suggesting that parts of the world constitute the domain of the function and the mathematical systems the range. In contrast van Fraassen appears to conceive the parts of the world as constituting the range of the mapping.

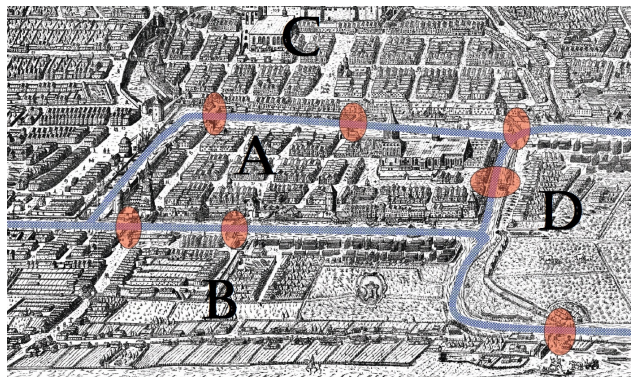


Figure 1.2.1: *Adapted Map of Königsberg, 1613.*

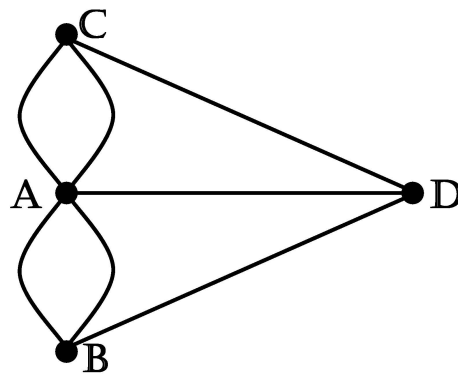


Figure 1.2.2: *Representation of Königsberg Bridge Structure.*

order to consider such a proposal, we must figure out what such “true but non-mathematical relations/structure” imposed on a physical system are supposed to be like.

To do this, let us consider one of Pincock’s more compelling example cases of mathematical representation, Leonhard Euler’s application of what is now called graph theory to investigate tours of the seven bridges of Königsberg. Euler states the problem as follows:

In the town of Königsberg, in Prussia there is an island  $A$  called “Kneiphof,” with the two branches of the river (Pregel) flowing around it, as shown in [figure 1.2.1]. There are seven bridges ... crossing the two branches. The question is whether a person can plan a walk in such a way that he will cross each of these bridges once but not more than once. I was told that while some denied the possibility of doing this and others were in doubt, there were none who maintained that it was actually possible. On the basis of the above I formulated the very general problem for myself: Given any configuration of the river and the branches into which it may divide, as well as any number of bridges, to determine whether or not it is possible to cross each bridge exactly once. (Euler, 1956, p574)

A contemporary treatment of this problem in the context of graph theory is to represent the Königsberg “bridge structure” as an undirected graph, in which land regions separated by water (e.g. the regions labeled  $A$ ,  $B$ ,  $C$ , and  $D$  in figure 1.2.1) are represented by vertices and the bridges connecting such regions are represented by edges joining the respective vertices. A quick comparison with figure 1.2.1 tells us that a graph representing the 1613 Königsberg

## 1.2. DEFENDING FIDELITY

bridge “structure” in this way can be symbolized as in figure 1.2.2.

In honor of Euler’s work on this problem, a sequence of vertices of a graph in which two vertices are listed sequentially only if there is an “unused” edge connecting them and each edge is used in this way exactly once is now suggestively referred to as an *Euler path*. If such an Euler path exists for a graph of the form of figure 1.2.2, then it is not hard to see how that sequence would constitute instructions for which bridges to take and in what order to take them in order to traverse the bridges in the way that reportedly eluded the Königsberg townsfolk. However, Euler’s theorem tells us that a necessary condition for the existence of such a sequence is that at most two of the vertices can have an odd number of edges joining it. So, since the Königsberg graph structure clearly fails this condition (all four vertices have an odd number of edges), we can conclude that no such Euler path exists for the graph in figure 1.2.2. So, since we concluded (by inspection of figure 1.2.1)<sup>19</sup> that the graph in figure 1.2.2 matches the actual bridge structure of the physical city of Königsberg (circa 1613), we can infer that it is impossible to traverse all the Königsberg bridges exactly once, just as the townsfolk suspected.

Of course, van Fraassen is absolutely correct that merely whistling at figure 1.2.1 and saying “this is the structure” does not count as providing a well defined mathematical system to which we can map the relations depicted by figure 1.2.2. Nonetheless, comparison of the two figures quickly cements the feeling that something homomorphism-like can be observed between the way that the bridges connect the four labeled regions of land, and the way that the edges of the graph connect its four vertices. Even if we abandon the terminology of ‘mapping’ and ‘morphisms’ is there a way to rehabilitate the motivation behind such an

---

<sup>19</sup>Of course the etching is itself a visual representation of the physical city in 1613. Because we have no way of directly verifying the properties of the city in the past we will have to make due with the assumption that the bridge features there depicted (along with other evidence like Euler’s own report in (Euler, 1956) etc.) are indeed veridical.

## 1.2. DEFENDING FIDELITY

account?<sup>20</sup>

In order to achieve such a rehabilitation, we must address the more significant issue of figuring out what the physical relations that a mapping theorist wishes to “homomorphically preserve” could be (even if it is not mathematical). First, we note that the (then not even nascent) mathematics of graph theory is actually unnecessary to draw the conclusion that such an Eulerian tour<sup>21</sup> across the Königsberg bridges does not exist. As Euler himself noted, since the number of bridges and land areas separated by the rivers is finite, a brute force strategy could be used to reveal that of all the potential walks across Königsberg bridges without repetition, none will include all seven. However, such a brute force strategy might seem impractical for arbitrarily large cities with arbitrarily large numbers of bridges.

Better yet, we might explain the general impossibility inferred from Euler’s theorem by considering a particular (entirely non-mathematical) argument about unique bridge walking.<sup>22</sup> We only need to make the *principle observation* that if the bridges to a particular region cannot be paired up into separate *in-and-out* pairs [i.e. if there is an uneven number of bridges] then it is impossible to walk across each bridge connecting that region without either getting “stuck inside” that region or becoming “stuck outside” of it with no bridges left to take. For example, if we start outside of Kneiphof (i.e. region *A* depicted in figure 1.2.1) and in the course of our attempt, travel into [1], then out of [2], eventually return into [3], and then leave again [4], and finally return [5], via the [five] separate bridges leading to Kneiphof, then we will ultimately be stuck on Kneiphof island, unable to leave via any of

---

<sup>20</sup>Contessa’s use of the ‘denotation’ terminology instead of ‘mapping’ terminology in his method of analytic interpretation (Contessa, 2007, pp57-8), might count as an example of such mitigation. That being said, even with this terminology, we are still faced with the question of how to determine the interrelations of the target that are supposed to be captured in the vehicle.

<sup>21</sup>To be precise, we will call the physical act of walking over the bridges in the city exactly once an Eulerian *tour*, to be contrasted with Eulerian *paths*, which refer to mathematical sequences of edges.

<sup>22</sup>The following observations are quite close to those used by Euler himself in his original demonstration.

## 1.2. DEFENDING FIDELITY

the bridges without using it for a second time. This works if we intend Kneiphof to be the final destination point, or (if we were to run the path in reverse) the starting point of our attempt, but there are only so many “end points” to a trip. More generally, it is ok to get “stuck” at a region if we want that region to be a starting point or destination point of an attempt to traverse all the bridges exactly once. The problem is that there is only one starting point and one destination point of any potential bridge tour of Königsberg. [In other words, we can only get “stuck” twice.] But our principle observation was that if a region has an unpairable [odd] amount of bridges, then we cannot traverse them all exactly once without getting “stuck.” Hence, for any city (including Königsberg), we can traverse every bridge in the city exactly once, only if we will have to get “stuck” in a single destination region and a single starting region. But, since our principle observation reveals that we have to get stuck not just in regions  $A$  and  $B$ , but also in region  $C$  and in region  $D$ , we conclude that there are not enough ends of the path to go around, confirming the townsfolk’s suspicions that such a tour of the bridges is impossible.

Though such an explanation of why the townsfolk’s suspicions were correct is somewhat cumbersome when completely purged (with brackets) of any mathematical language, it is difficult to argue that such an explanation is a case of mathematical application or representation. This is not to say that the math does not make it much easier to communicate the explanation clearly and concisely, at least to those who are a bit familiar with the terminology. The point is only that the mathematics itself did not seem to be necessary for the expurgated version. On the other hand, certain presuppositions were absolutely vital both to the graph theoretical explanation as well as the math-purged version. We had to assume constraints on walking the Königsberg bridges in order to justify the impossibility of an Eulerian tour. What are these constraints?



## 1.2. DEFENDING FIDELITY

There are in fact a number of ways to traverse all seven bridges exactly once. If an eighth pathway between any of the two regions were to exist, then one could easily take an Euler tour of all the bridges plus the additional path. Now, imagine a mischievous agent bringing a temporary rope bridge along on the tour, pausing part way through the tour to wait for winter when the Pregel river freezes over, inventing a flying contraption, arranging for a boat at some point, or any number of other possibilities. Such an agent could (in a sense) foil the Königsberg townsfolk’s suspicions.

This does not mean that Euler’s proof was invalid. It also does not mean that the math-purged explanation given above is faulty. We might say that by using such strategies our mischievous agent has cheated. She broke the ground rules that we assumed governed the way the problem was set up to begin with. And these ground rules were necessary in order to generate the above arguments (math or no math) for why such bridge tours are impossible. To take advantage of temporary or non-bridge pathways across the river would violate our so-called “principle observation” in the expurgated case. Similarly, such mischievous cheating would falsify the claim that the available paths are those depicted by the structure of figure 1.2.2. In both cases, in order to draw the inference that Eulerian tours are impossible, we needed to assume, for instance, that the bridges are the only (legitimate) method of river crossing and that the “bridge structure” has a certain *fixed* configuration.<sup>23</sup> Moreover, we had to presume a whole host of other conditions on what would and would not count as unique bridge crossing (like walking along the left side of a bridge is equivalent to walking along the right side or down the middle, if you close your eyes, hop on one foot, or someone carries you, it still counts, and so forth).

---

<sup>23</sup>As an historical irony, during World War II two of the seven bridges of Königsberg (now called Kaliningrad) were destroyed and never replaced, making it now possible to take an Eulerian tour of the remaining bridges.

## 1.2. DEFENDING FIDELITY

Such constraints are so painfully obvious in the Königsberg example that we hardly notice. This is part of what makes the example so compelling. It takes little work to see that figure 1.2.2 captures all and only the “legitimate” bridge relations depicted by the etching of figure 1.2.1, rightly treating as equivalent insignificant variations in how one might “uniquely” walk across one of the bridges as well as eliminating other kinds of tricks a mischievous cheater could pull. In short, figure 1.2.2 provides a concise summary of the bridge relations that we might already abstract from inspection of the etching. Again, abstracting in this way is so natural in the Königsberg example that we hardly notice that we’ve done it. However, as exemplified by the mischievous cheater, without this process of laying out the initial constraints on the physical system in question (Königsberg circa 1613) it is impossible to answer Euler’s question, whether we use the math or not.

Despite the already emphasized simplicity of the Königsberg example, the process of abstraction is not a straightforward matter of simply observing the physical existence of the bridges and where they are located. Abstraction processes involve the non-algorithmic skill of identifying which specific features of the bridges are relevant while judiciously ignoring other details that do not matter. In this case, the most relevant features include the facts that one can travel both ways on any of the seven bridges (they are undirected), that more than one bridge can be connected to a particular region, that once you are at one end of the bridge you can access those (and only those) bridges connected to that side, and, of course, how each of the bridges connects the respective four regions separated by the river. Features of the bridges that are irrelevant include, differences in their material construction, where along banks of the river branches they have been built, variations in length and width, perturbations in methods of getting across a particular bridge, etc. The process of abstraction involves culling through all of these ostensibly available properties and interrelations of the

## 1.2. DEFENDING FIDELITY

components of the entire physical system (i.e. Königsberg in 1613) and picking out the specific features that are germane while eliminating those that are not.

Once an abstraction has taken place, van Fraassen’s category error challenge becomes nominal at best. Unlike before, after an abstraction has been conducted, we have a well defined system of relations which can (if we wish) be “homomorphically”<sup>24</sup> preserved through some correspondence with a mathematical system. Hence, particularly in the case of scientific representations of physical phenomena, if we have a way to abstract all and only the interrelations of a physical system that we wish to represent with our theory, we would then be in the right position to employ a homomorphism-like mapping strategy whereby the abstracted interrelations of the physical system are analogously preserved by the strictly defined constraints placed on the structure of some mathematical vehicle. In other words, once we have abstracted relevant physical relations from the system, applying the math in a non-mysterious way ceases to be a problem, because the physical relations can then be simulated in the mathematics. If we have correctly abstracted a well-defined structure from the physical system, then the faithfulness of inferences made by means of a homomorphically analogous mathematical system will no longer be mysterious. The surrogate inference can then be justified as sound thanks to the homomorphic analogy.

---

<sup>24</sup>It is debatable whether the relations so abstracted themselves properly count as mathematical or simply abstract. For those inclined to the latter interpretation, it might still seem inappropriate to refer to such relation preserving correspondences between the abstracted physical components and a strictly defined mathematical system a “homomorphic mapping,” since such terminology is often used in more exclusively mathematical contexts. That being said, in cases where the physical relations are already abstracted and hence a well defined relational structure is available, we no longer are faced with the substance of van Fraassen’s challenge (i.e. that there is not a well defined target to “map” to). Moreover, the literal meaning of the Greek origins of the term ‘homomorphic’ give us reason to use it in this broader employment. In the following we will hence continue to use mapping terms like ‘homomorphism’ and other variants with the explicit caveat that whether or not the narrow mathematical usage is intended will be dependent on the context.

## 1.3 Ampliative Abstraction and Epistemic Debt

We now have a potential defense against the problem of mysterious fidelity. As we have seen mapping accounts identify a potential justification of how certain mathematical representations become faithful by appealing to the existence of a kind of analogy of relevant interrelations of the parts of the mathematical vehicle of representation and its physical target. The typical mapping proponent suggests appealing to some sort of functional mapping between the parts of the physical target and the parts of the mathematical vehicle that will establish the right kind of analogy, such as a homomorphism or isomorphism. The problem with this solution (raised coyly by van Fraassen's category error challenge) was that in the case of the physical target, it is not predefined what the relevant parts of the physical system are supposed to be and which interrelations are supposed to be preserved by such an ostensibly homomorphic mapping. To solve this challenge, we observed in the last section that for successful mathematical applications through representation what we need is a procedure of abstraction whereby the representer precisely identifies specific relevant components of the physical system to be represented and their interrelations in order to provide a set of well defined abstract physical relations which can then be homomorphically preserved in the mathematical vehicle of representation. If such an abstraction procedure is both available and successful in capturing actual and relevant interrelations of the physical system to be represented, the mysterious fidelity problem will be resolved. So, since our primary concern is with the mathematical representation of physical phenomena in science, we must now take a careful look at how such abstraction procedures come about in scientific practice.

There is little question that when it comes to the representation of physical phenomena in science, an essential component to determining the relevant interrelations of the physical

### 1.3. AMPLIATIVE ABSTRACTION AND EPISTEMIC DEBT

system is observation. The difficulty lies in the point already hinted at in the last section, namely, that, though observation is no doubt central and vital, the abstraction of interrelations from observation alone is far from apodictic. In this section we will take a careful look at how abstracted interrelations represented by mathematical vehicles of representation can arise in scientific practice. In particular we will emphasize two significant stages of ampliative inferences involved in developing such abstractions, which we will have to address in developing our account of how matching might be sufficient for unmysteriously successful scientific representation in chapter 2.

#### 1.3.1 The Observation of Data vs. Phenomena

In order to understand how physicists go about abstracting the relevant features of a physical system from their observations of the system (predominantly by means of experimentation), we must be able to answer the question of what it is that they actually observe. To answer this question it will be useful to take advantage of the now seminal distinction between *data* and *phenomena* drawn by James Bogen and James Woodward in their (Bogen & Woodward, 1988). The concept of *data* in this terminology consists of the “records of effects in investigators’ sensory systems or experimental equipment” (Bogen, 2011, p8). It is constituted by unmanipulated, unprocessed experimental results of the investigation of the physical system or system type in question. Examples of data may include, thermometer readings, reaction times, discharges on a particle detector and so forth. Data can include irrelevant “noise” resulting from particular idiosyncrasies of its extraction, including statistical fluctuation, impurities in the object of measurement, imperfections on particular measurement occasions or instabilities in the apparatus.

Data plays the role of supplying evidence for the phenomena being studied. In this

### 1.3. AMPLIATIVE ABSTRACTION AND EPISTEMIC DEBT

terminology the *phenomena* investigated is characterized by the “stable, repeatable characteristics” of the physical system or system type (Bogen & Woodward, 1988, p117). So, for instance, in one of their paradigm examples of investigating the phase transitions of lead, the particular thermometer recordings taken upon the melting of a number of individual lead samples (in otherwise *ceteris paribus* circumstances) would count as the experimental data, while “the” melting point of lead (identified perhaps through some aggregation of these entries, say, by averaging) would count as the studied phenomena of the experiment. An important feature to note (and exemplified by this case) is that phenomena are not always best thought of as being instantiated (at least not in any straightforward way) by a particular experimental occasion. For instance, in the lead example none of the particular samples may actually result in a recording of exactly  $327.5^{\circ}C$  even though this aggregation number is supposed to report the detected phenomena.<sup>25</sup> Instead, phenomena are perhaps better thought of in terms of the *stable patterns* present across a number of equivalently prepared systems or experimental occasions.

Bogen and Woodward are also careful to emphasize that while “[p]henomena are detected through the use of data,” they are typically “not observable in any interesting sense of that term” (Bogen & Woodward, 1988, p306). Though phenomena are indirectly detected by our direct observation of the data (along with various inferential judgments that such data counts as evidence of the phenomena in question), *phenomena are not directly observable*. That being said, the primary kind of “unobservability” of phenomena at issue must be carefully distinguished from another sense of unobservability often discussed by philosophers of science, particularly in the context of scientific realism. One way to be unobservable is in the sense

---

<sup>25</sup>As we shall elaborate in chapter 2, this number does indicate what can be thought of as the center of a kind of “neighborhood” in which lead samples in general will begin to change phase, but falling in such a “neighborhood” is still not a straightforward case of the sample instantiating lead melting at exactly  $327.5^{\circ}C$ .

### 1.3. AMPLIATIVE ABSTRACTION AND EPISTEMIC DEBT

that, say, neutrinos are unobservable. Neutrinos count as unobservable in that (according to our theories) they are far too tiny to see. They are beyond the limited scope of our sensory capabilities as human agents. It is tempting to characterize the unobservability of phenomena in terms of this sort of unobservability, thinking of phenomena as being unobservable because they are, so to speak, beyond the veil of our limited sensory capabilities, and therefore their detection must be mediated by data recovering experimental devices which we *can* immediately perceive. Though this might also be the case for certain phenomena, it is not the only (or even primary) sense in which phenomena are unobservable. The primary unobservability of phenomena lies in the fact that phenomena unlike neutrinos are not objects or entities, instead they are characterized by *the stable patterns* exhibited by physical systems of a certain type.<sup>26</sup> So, since the features of even *macroscopic patterns* cannot be directly observed in a single instance, we can say that phenomena are unobservable in the sense that no pattern is directly observable (at least not on a particular occasion).

The unobservability of phenomena is a Humean point: Just as we cannot directly observe a piece of bread nourishing, we cannot directly observe phenomena. Of course, with the help of antecedent theoretical presuppositions we may well infer from a number of instances that bread nourishes, and having warmed up a number of samples of lead (at constant pressure etc.) we may infer from thermometer data that lead melts at  $327.5^{\circ}C$ . Moreover, once we have concluded that such a phenomena occurs in these types of instances, we can easily recognize (detect) that a new piece of lead is melting (at least approximately) at  $327.5^{\circ}C$  in

---

<sup>26</sup>When claiming that phenomena are not entities, it is not being claimed that phenomena cannot be indicative (at least to a scientific realist) of the existence of an entity or event kind in the relevant physical system(s). For instance, reflection on various electromagnetic phenomena might lead a physicist (or realistically inclined philosopher of science) to infer the existence of a theoretically postulated entity called an electron, or reflection on photoelectric phenomena might lead a physicist to infer the existence of photons, etc. The theoretical postulation of such entities explains the presence of the phenomena for which data recovered in experimental tests provide evidence. But each of these three (entities, phenomena, and data) plays a distinct role and falls under respectively distinct categories.

### 1.3. AMPLIATIVE ABSTRACTION AND EPISTEMIC DEBT

accordance with the phenomena. *However, such recognitions are still not cases of directly observing the phenomena because in order to make such a recognition we had to already presuppose that the phenomenal pattern in question occurs in that type of instance.*<sup>27</sup>

Last, despite this difference in the observability of data and phenomena, Bogen and Woodward frequently emphasize the fact that whereas “data typically cannot be predicted or systematically explained by theory” due to inherent fluctuations of collection, “well-developed scientific theories do predict and explain facts about phenomena” (Bogen & Woodward, 1988, pp305-6). With this final point in mind, by adopting their terminology, we have a clear answer to the question of what it is that physicists directly observe, the data. On the other

---

<sup>27</sup>The stark distinction identified here between the directly observable (data) and the exclusively indirectly detectable (phenomena) drawn in the context of scientific theorizing might potentially raise the eyebrows of readers familiar with the arguments against sense-data theories of knowledge. Reflecting on the famous example of Sellars’s tie merchant (see (Sellars, 1963, pp142-6)), who has become so familiar with the way blue ties look under new lighting that he can now immediately “observe that it is blue” even though it appears green, we might wonder how it can be possible to immediately see (after a number of instances) that the tie is blue in this lighting, while according to the data-phenomena distinction it is impossible to eventually see that the melting point of lead is  $327^{\circ}C$ . After all, one point of the tie example is to disabuse the notion that there is a firewall between certain contents that can directly be seen (viz “the given” or sense data) and certain facts that cannot be. Hence, one might worry that we have somehow violated this principle by suggesting that experimental data have been distinguished from phenomena in just this way.

A few responses can be provided to those with these concerns: First, we must remember that part of the problem with sense data theory, leading to Sellars’s rejection, was that it in fact tries to treat perception in the model of scientific theories, in which sense data indeed play the role of data and facts about the world play the role of the phenomena. Since in contrast we are restricting our distinction exclusively to the context of scientific investigation (not the so-called “manifest image”), the distinction between data and phenomena does not succumb to the myth of the given. Moreover, as emphasized above, there is an important difference between say “seeing the phenomena of gravitational attraction between two massive bodies” and “seeing that two bodies are approaching each other *in accordance with the phenomena of gravitational attraction*” the former is unobservable in the sense we have described, whereas (if we are familiar enough with gravity) the latter is immediately and directly observable on analogy with how the tie merchant can directly observe the blueness of the tie (having become familiar with the fact that blue ties look green in the new light). Again, recognizing in such a case that it is the type of situation in which gravitational phenomena occur and that its occurrence accounts for what we have observed (the approach of the bodies) is different from seeing the pattern itself on such an occasion. Finally, by drawing the direct observability distinction between data and phenomena we by no means must preclude the possibility of, so to speak, “seeing a pattern in the data.” Indeed, an experienced experimental physicist may very well possess such an ability and this kind of skill can be an asset in successful abstraction. However, to immediately “see/perceive/observe” patterns in data is quite different from observing a phenomenal pattern directly on a particular experimental instance. In the former, in order to “see” the pattern one must inspect the spectrum of data collected whereas in the latter, one is only inspecting a particular instance, and so for the reasons already given cannot be observing the (entire) phenomenal pattern directly.



### 1.3. AMPLIATIVE ABSTRACTION AND EPISTEMIC DEBT

hand, it is the properties and interrelations attributed to the phenomena that seem to be the sort of features from which we would expect to reliably draw inferences. Though we have direct observational access to the results of measurement devices in our experiments, the presence of experimental noise and other fluctuations that are included in data make such raw measurement results a poor candidate for the abstracted interrelations to be captured by means of a mathematical representation. Though unreliably fluctuating elements are bound to occur in most empirical investigations, such fluctuations are not the kind of features that we want to preserve and use as a basis for drawing (surrogate) inferences about systems of the relevant type. In contrast, the stable and repeatable patterns characteristic of phenomena are exactly the kind of desired qualities we might want as our bases for inference about physical systems. This thesis is punctuated with the fact emphasized by Bogen and Woodward that phenomena, not data, are the proper target of theoretical prediction and explanation. That is to say, *surrogate scientific inferences are made about phenomena not data*.

So, while the stable interrelations attributed to the phenomena serve as the appropriate kind of candidate for “physical structure” to be preserved in faithful representations, this structure is not what is directly observable by means of experimentation. We have hence identified an observability gap between “directly observable” data records on the one hand, and on the other hand the interrelations attributed to the phenomena which fail to be directly observable. Since it is the latter interrelations we actually want for our faithful representations, the bridging of this gap constitutes a significant ampliative inference in actual scientific abstraction practice, which we will refer to as the *data to phenomena gap*. In addition to this data to phenomena gap, there is a second gap between the system and the supposedly “directly observable” data records. In sections 1.3.2 and 1.3.3 below we will review some of the non-apodictic features of how data is recovered from physical systems

through experimentation, and in section 1.3.4 we will then review the non-apodictic features of how the data to phenomena gap is bridged in turn.

To be clear, these sections should not be taken as a critique of scientific practice. These sections instead aim to emphasize that the abstraction practices actually used by scientists to recover well determined interrelations possessed by physical systems is not a straightforward or deductive task. The abstraction of phenomenal patterns in physical systems is inferentially ampliative, and as we shall see highly sensitive to the adoption (and revision) of background empirical and theoretical presuppositions. We will return to this premise in section 2.2, where in resolving what we will term as *Plato's problem* we reconsider what physical phenomena are targeted by mathematical idealizations in a representation relationship.

#### 1.3.2 Determining the Data: Rejecting Naivety

According to the account just laid out data records are recovered from experimentation and these records serve as evidence for the detection of scientific phenomena. The detection of these phenomena can in turn be used as the basis for (an abductive) justification of scientific theory, explaining their detection in the experiment. Though such a “linear” progression captures the principle stages of the process Bogen and Woodward propose, endorsing their account is not to suggest a naive view of scientific investigation in which the initial data recovery of experimentation occurs in a theoretical vacuum.<sup>28</sup> According to such a naive account, theory construction develops directly from the bare perception-determined data without any presuppositions about the physical system type being measured.

Initial difficulty with such a naive account begins with the fact that data cannot be

---

<sup>28</sup>Both Bogen and Woodward have moved to reject this naive attribution in recent reviews of their data-phenomena distinction (see (Bogen, 2011, Woodward, 2011)), along with rejecting the idea that theory cannot inform “data to phenomena reasoning.”

### 1.3. AMPLIATIVE ABSTRACTION AND EPISTEMIC DEBT

determined merely (so to speak) by “whatever is sensed” in the course of the experiments. In dealing with a similar task to our own of accounting for how mathematics “coordinates” with physical systems Hans Reichenbach emphasizes the complications arising from identifying data with the bare perception of experimental apparati as follows:<sup>29</sup>

The content of every perception is far too complex to serve as an element of coordination. For instance, if we interpreted the perception of the pointer of the manometer... as such an element, we would get into difficulties because this perception contains much more than the position of the pointer. Should the factory label be on the manometer, it would be part of the perception. Two perceptions different with respect to this label may still be equivalent for the coordination to Boyle’s equation. Before a perception is coordinated, its relevant components must be distinguished from the irrelevant ones; that is, it must be ordered. (Reichenbach, 1965, pp40-1)

Reichenbach’s description of how the physicist must carefully cull the “relevant components” from “irrelevant ones” in our total perception of the state of the manometer is strikingly similar to our discussion of the task of identifying the relevant features of the Königsberg bridges at the close of section 1.2.3. Before we can even identify particular observations of experimental results as potential evidence of some phenomenal pattern suitable for mathematical representation, the physicist must decide first how to conduct a measurement and second what properties of the result of the measurement are supposed to be recorded as data and what properties are inconsequential.

Of course, when it comes to recording the relevant results of a pressure measurement conducted with a manometer, the suggestion that we don’t have a clue as to what to record might seem rather obtuse. We know the relevant result to record: as indicated by Re-

---

<sup>29</sup>In many ways, Reichenbach’s program of coordination can be thought of as an early twentieth century attempt to deal with the very issues of representing physical systems with mathematics that we have been investigating. He too identifies and attempts to solve the dilemma posed in the introduction to this chapter as well as the particular challenges, such as van Fraassen’s category error challenge, associated with “mapping-like” attempts (see (Reichenbach, 1965, pp34-9)). Reichenbach’s solution in (Reichenbach, 1965) of relying only on consistency of coordination is unfortunately shaded by the perhaps overly empiricistic scruples he held at that time, but the major dilemmas he identifies are still remarkably appropriate (cf. (van Fraassen, 2008, pp218-23)).

### 1.3. AMPLIATIVE ABSTRACTION AND EPISTEMIC DEBT

ichenbach himself, it is “the position of the pointer.” Once we have chosen well accepted measurement apparati and techniques of application, the answer of what to record for the data is manifest. But the development of a well accepted measurement apparatus already involves a great deal of antecedent theorizing. This antecedent theorizing informs (a) how to construct the relevant apparatus, (b) our ability to conceptualize what properties or relation of properties are measured, (c) why we expect that the apparatus is able to indicate the properties so conceived, and (d) how to display the relations in the data output of the apparatus. Without at least some antecedent knowledge or belief allowing an experimentalist to answer these requirements, she has no hope of even gathering data (let alone guidance in conducting the sophisticated task of inferring facts about reliable phenomenal patterns from such data records). How the scientist answers these requirements comes by way of a protracted and recursive process whereby immature methods of data collection are informed by immature theorizing (or proto-theorizing) which leads to (ampliative) judgments about phenomena, allowing for improved theorizing which in turn informs data collection methods once again, and so on. In the next subsection, we will briefly review a concrete (albeit somewhat apocryphal) example of this recursive process, taking a look at the well discussed case history of temperature measurement.<sup>30</sup> As for our primary task, the modest point to appreciate has already been made: *contrary to the naive view, even the recovery of our observable data cannot be conducted without significant (theoretical or proto-theoretical) presupposition concerning the type of system studied.* The appeal to such antecedent supposition hence marks an ampliative gap between bare interaction with physical systems and the recovery of data records. This gap is bridged in actual scientific practice only through significant

---

<sup>30</sup>For more extensive critical discussion of the development of temperature measurement in this context see (Mach, 1986, pp10-61) and (Chang, 2004). See also (van Fraassen, 2008, pp125-30).

ampliative reasoning used to support the required antecedent theoretical supposition.

#### 1.3.3 Determining the Data: Temperature

Early devices for measuring temperature changes, frequently attributed to Galileo for their invention, consisted of a long inverted flask with the open end submerged in a small glass basin of water. The idea was that observations of the volume of the water in the flask might serve as an indication of temperature. Retrospectively, we can expect that on warm days Galileo would observe a large volume of water rising up into the flask as the water dilated, whereas on cooler days the volume would decrease. Similarly, if the basin were to be warmed, say by one's hands, Galileo might have observed the water in the flask rising, and then falling again some time after the basin had been set down.

Now imagine it is the turn of the seventeenth century and we have joined Galileo for the development of his thermoscope.<sup>31</sup> He is faced with the challenge of figuring out how to record such potential “thermal data.” One suggestion we could give is to make some kind of markings along the side of the flask and label each marking with a different name, say, arbitrarily chosen names of prior Popes. These markings will then enable him to record, for instance, that in such and such circumstance the water in the flask had risen to *Leo X* whereas under some different circumstance, the water had risen only to *Paul II*. Galileo's thermal measurement device now (ostensibly) allows him to record rudimentary comparisons of the relative warmth or coolness of different circumstances. By reviewing his notes of which names were recorded in which circumstances, he might judge (from inspection of where the pope name was along the device) that on occasion *A* there was more/less warmth than on

---

<sup>31</sup>We will follow the convention of referring to devices capable of (potentially) measuring changes in temperature as *thermoscopes*. In contrast, we will refer to devices capable of (potentially) measuring temperature quantitatively as *thermometers*.

### 1.3. AMPLIATIVE ABSTRACTION AND EPISTEMIC DEBT

occasion *B*. To avoid the need for constant reference to how he arbitrarily listed the pope names along his device, Galileo could even improve his recording system, by relabeling the markings so as to take advantage of a natural ordering already well known to exist between past popes, namely, their order of ascendancy. He can relabel the markings by naming the top most marking *Clement VIII* (the likely current pope during his invention of the first thermoscope), and working his way back naming the next highest marking with the name of the next most recent pope, and so on. By taking advantage of this ordering “structure” of when each pope presided, he can now impose such an ordering on his recorded data. If the pope name recorded on occasion *A* is of a pope who presided after the pope whose name was recorded on occasion *B*, Galileo might surmise that it was warmer on occasion *A*.

Before improving further on Galileo’s thermoscope, note the significant ampliative steps that have already been implicitly taken even by imposing such a meager ordering. Galileo’s thermometer is designed on the notion that comparisons of warmth can be made by comparisons of the volume of the liquid that has risen in the flask. Presumably this presupposition could be confirmed by the experiences already noted of the water rising on warm days and falling on cool ones. Though *significant* temperature differences might be evident for *significant* changes in water level, such experiences do not entail that this association (the warmer it is the higher the water) is preserved for every potential difference in warmth. Just because we can observe that very warm days result in very high water and very cool days result in very low water, this does not (deductively) entail that subtle variations in warmth result in analogous variations in water level. That is to say, he has made an ampliative inference that the association between water level height and warmth exhibited in severe cases generalizes to cases where the differences in warmth are quite small (at least to the differences as small as he has spaced the “pope markings”).

### 1.3. AMPLIATIVE ABSTRACTION AND EPISTEMIC DEBT

From a contemporary perspective, a skeptic might challenge such an ampliative inference by pointing out that though we observe through “direct” experience that it *feels* cooler when our (contemporary) thermometer displays  $15^{\circ}C$  then when it displays  $32^{\circ}C$ , this does not deductively entail that it is analogously cooler when our thermometer displays  $25.00^{\circ}C$  than when it displays  $25.01^{\circ}C$ .<sup>32</sup> It is well documented that such variations in temperature are below the threshold of human sensation. Moreover, it had been observed long before Galileo’s (supposed) development of early thermoscopes that human assessment of temperature comparisons are notoriously unreliable.<sup>33</sup> This is not to say that the judgment that a total ordering can be given to temperature records is in error, only that drawing such a judgment involves significant ampliative inference in light of the stark absence of reliable ordering capacity possessed by humans.

Other potential ampliative judgments regarding Galileo’s open air thermometer were eventually determined to be unwarranted. For instance, we might note from a contemporary perspective, that variations in atmospheric pressure can introduce a significant confounding influence on the Galileo’s data results. Though “pope name” recordings may be reliable in a spatially and temporally local sense, changes in atmospheric pressure from day to day or altitude to altitude might significantly influence such measurement. But how is Galileo even to discover such confounding influences, particularly since barometric pressure measurement practices were also yet to be fully developed? It would have been possible, if he had some sort of *fixed point* of temperature. That is, we could determine that an influence such as

---

<sup>32</sup>This sort of “skeptical” challenge is not entirely gratuitous. As we know now, the ampliative generalization in question fails to hold for every choice of liquid and every temperature variation. For instance, water reaches its maximum density around  $4^{\circ}C$  and then begins to expand again as it approaches  $0^{\circ}C$ .

<sup>33</sup>Consider Berkeley’s famous experiment in which an agent sticks each of her hands in respectively cool and warm buckets of water and then places them into a third bucket of water of medium warmth. Recognition of this unreliability has been documented well into antiquity (for one of the earliest such hints see (Plato, 1997, 154b)).

### 1.3. AMPLIATIVE ABSTRACTION AND EPISTEMIC DEBT

atmospheric pressure is confounding our results by identifying a kind of thermal state which always occurs at *the same temperature*. Unfortunately, this task was far from straightforward and fraught with false starts. Candidates for fixed points (which can appear somewhat foolish from a contemporary perspective) included candle flame (Sanctorius), the “greatest summer heat” (Accademia del Cimento), melting butter and deep cellars (Joachim Dalencé), and body temperature (Issac Newton)(Chang, 2004, p10). The process of ruling out these along with various other candidates again involves ampliative judgment and supposition. For instance, we might expect that multiple humans in the same spatial location at the same time should have the same body temperature if it is to be a fixed point, so we can rule out Newton’s candidate by comparing a number of human body temperatures and judging that there is significant variation among the results. Unfortunately, in order to determine that the variation among results was *significant* we must first have an acceptable method of determining the data and second have a method of analyzing whether variation is indeed significant or was actually indicative of a reliable “fixed point” phenomena (i.e. *the* human body temperature). But the reason we wanted a fixed point in the first place was to help us achieve the former task.<sup>34</sup> Moreover, as we shall discuss in section 1.3.4 such data to phenomena reasoning required for the second task itself involves a good deal of ampliative reasoning.

Fixed point standards did not become fully settled until well after Galileo’s death. However, the parallel development of barometric measurement and theoretical conceptualization eventually led Blaise Pascal to identify the influence of atmospheric pressure on open air thermometers (van Fraassen, 2008, p126). This resulted in the development of closed or

---

<sup>34</sup>From a retrospective vantage, we expect Galileo’s open air thermoscope to be locally reliable because the confounding influences of atmospheric pressure are locally stable, but if we already knew that this was a source of error, we wouldn’t need a fixed point to tell us. *Cf.* Chang’s poetical comparison of this challenge with the task of hanging a picture on a wall that hasn’t yet been built. (Chang, 2004, p40)



### 1.3. AMPLIATIVE ABSTRACTION AND EPISTEMIC DEBT

“liquid” thermometers where a heated liquid was introduced into a glass tube that was then sealed. Initially numerous different liquids were used in such thermometers. The problem with such lack of standard liquid choice was that different liquids expand at different ratios with changes in warmth. Moreover, the ratio rates are not constant as warmth (temperature) varies. Such facts suggested that perhaps a standard liquid convention would need to be chosen for a particular temperature measurement scale.<sup>35</sup>

With a convention of standard liquid choice accepted (by ampliative argument) to react to thermal influences similarly from sample to sample as well as conventional enclosure (e.g. glass cylindrical tube), we can again consider how the potential “thermal data” measured by such thermoscopes can be recorded. Our fictitious suggestion that Galileo use pope names for his thermoscope markings would be useful in detecting changes in warmth, but fails to establish a measurement of temperature level that can be recorded and then compared to warmth states in wildly diverse situations. Reflecting on this point Ernest Mach muses that such standardized liquid thermoscopes could at least have been used to detect the proper marking levels of thermal fixed points. After making the ampliative inference that such a standardized thermoscope indeed detects such fixed points reliably under arbitrary variations of other conditions,<sup>36</sup> he suggests that such a thermoscope could indeed be used to detect differences in temperature as desired. Mach immediately reviews the manifest difficulties

---

<sup>35</sup>Choice of a standard liquid, such as mercury, involves its own ampliative generalizations. Though it was observed, for instance, by Dalton that mercury “appeared to have the least variation” (see (Mach, 1986, p54)), it still evidently varied in ratio compared to other liquids. Moreover, glass expands roughly at a ratio of seven to one with respect to mercury, meaning that the marking system is not static with temperature variation either. On both counts, we are hence forced either to generalize and make the ampliative judgment that such variations are negligible or to attribute a rather uncomfortable position of prominence to our conventional choices of glass and mercury, insisting that thermal states are inextricably pinned to “whatever the mercury-glass thermometer does in that state.”

<sup>36</sup>Again this is not a trivial conclusion to draw. It requires both that such fixed points exist under arbitrary variation (they don’t), and (as already discussed) that we can presume a linear ordering to the temperatures at which these fixed points reliably occur. For further discussion of the difficulties with identifying such fixed points see (Chang, 2004, ch1)).

### 1.3. AMPLIATIVE ABSTRACTION AND EPISTEMIC DEBT

and limitations of such a purely qualitative standardized thermoscope:

The inconveniences of such a system, which as a matter of fact long prevailed would soon be manifest. The more delicate the inquiry, the more fixed points of this sort would be necessary; and ultimately they would no longer be attainable. Furthermore, the number of the names to be marked would be annoyingly augmented, and it would be impossible to discover from these names the order in which the thermal states under consideration succeeded one another. (Mach, 1986, pp49-50)

Like the above discussed fictitious pope markings system, even with such a standardized thermoscope, the practice of using otherwise unordered fixed points has an impact on the practical comparison of data recordings (unlike popes, fixed points don't have an independent antecedent ordering). More importantly, a nominal system of merely labeling the fixed points only allows for the qualitative comparison of various states. Under the presumption that there is a total ordering that can be imposed on temperature levels, the best such a thermoscope could do is identify that a given measured state is “between” two previously identified fixed points. Mach hence proceeds to note the obvious availability of “a system of names which is at the same time a system of ordinal symbols, permitting of indefinite extension and refinement, viz, numbers” (Mach, 1986, p50). The benefits of using a “system of names” with such a well known ordering structure are clear: Unlike our pope names system, by using the real (or rational) numbers, the developer of the thermometer could make use of the density and unboundedness of their ordering to compare arbitrary disparate variations of thermal states (there is no upper bound to potential temperature measurements recorded by real numbers) and to arrange them with arbitrary precision (between any two temperature measurements, there can exist a record of an intermediate state). However, by choosing to use such a richly structured “system of names,” the thermometer creator presupposes that the actual thermal states possible admit such arbitrary precision and unboundedness. But, as was already evident from our pope ordering, such a presupposition requires inductive

### 1.3. AMPLIATIVE ABSTRACTION AND EPISTEMIC DEBT

generalization.

Even if it were the case that we could observe an invariant ordering of fixed points with a standardized thermoscope, as already noted, this does not deductively entail that there *is* an invariant ordering of *all* thermal states. How do we justify that a state measured by a (contemporary) thermometer at  $23.0001^{\circ}\text{C}$  is “really warmer” than a state measured at  $23.0000^{\circ}\text{C}$ . Further, how do we justify that thermal states *should* be differentiated with arbitrary precision? There is reason to believe that there is a “thermal” difference between  $100^{\circ}\text{C}$  and  $0^{\circ}\text{C}$  because of the difference between the boiling and melting of water, but if there are no fixed points to distinguish between a situation where the thermometer gives a reading of  $23.0001^{\circ}\text{C}$  and a situation where it gives a reading of  $23.0000^{\circ}\text{C}$ , it does not automatically follow there is a thermal difference between the two (even if there is a detectable difference in the volume of the thermometer liquid). Imposing such a richly structured “system of names” hence requires the ampliative judgment that such a structure exists for the thermal states being measured. Choosing to use the real (or rational) numbers to record these differences suggests that such differences can exist. But the judgment that such differences do exist cannot be drawn deductively from our knowledge of more coarsely differentiated thermal states. In other words, the only answer to how we might justify totally ordering thermal states with arbitrary precision is by ampliative generalization (from our knowledge of the coarsely differentiated thermal states).

Again, by recognizing that such ampliative inference is inherent, especially when we employ numbers for our measurement records, we are not proposing a thesis of skepticism. Such a move may well be epistemically warranted. The point is only to note, that by appealing to such a richly structured “system of names” like the real or rational numbers we must (so to speak) fill in gaps that cannot be justified by deduction and observation

### 1.3. AMPLIATIVE ABSTRACTION AND EPISTEMIC DEBT

alone. By filling in those gaps by presupposing that a such a richly structured measurement system is appropriate, we are going beyond what is immediately justified by prior experience and experimentation alone. Using rich mathematical systems to record the data as in the example of temperature measurement, hence, constitutes a significant ampliative move that must be taken in actual scientific practice. The presuppositions required for such ampliative judgments constitute an epistemological voucher: *Even at the “data extraction” stage of abstraction, the physicist must “borrow” on the presupposition of structure (in this case the metric structure of  $\mathbb{R}$ ) that is not directly warranted by experimentation.* In chapter 2 we will argue that such epistemic debts can be rebalanced (in part) through our proposed account of scientific representation.

#### 1.3.4 Data to Phenomena Reasoning

In the last two subsections we have argued for the existence of an inferential gap between knowledge we can gain by direct interaction with physical systems and data records recoverable through experimentation. As we saw in the case history of thermometry, ampliative inference is inherent, particularly if we wish to record the data by means of richly structured recording systems such as the real numbers. This move required the presupposition of properties of thermal states (e.g. between any two temperature states there exists an intermediate state) for which there exists no deductive justification. Hence, we suggested that such abstraction judgments may incur what we characterized as a kind of “epistemic debt.” Let us call the inferential gap that must be bridged by incurring such a debt the *system to data gap*.

As already noted at the close of section 1.3.1, a second gap that must be bridged exists in moving from the raw data results of experimentation to claims about the phenomenal

### 1.3. AMPLIATIVE ABSTRACTION AND EPISTEMIC DEBT

patterns exhibited by the physical systems studied. Since we also noted there that such phenomenal patterns capture the kind of stable properties and interrelations that could be used for sound surrogative inference about the systems themselves, we must hence find a way to bridge this second gap as well. As with the system to data gap, the gap between data and phenomena must also primarily be bridged by ampliative inference.

In his recent review of the data-phenomena distinction, Woodward offers an account of how the data to phenomena gap is rationally bridged. He suggests that “data to phenomena reasoning” is “ampliative in the sense that the conclusion reached (a claim about phenomena) goes beyond or has additional content besides the evidence on which it is based (data)” (Woodward, 2011, p172). Though the literature on philosophical accounts of *ampliative inference* (or *reasoning*) is vast, the concept typically refers to inferences where the conclusion is not deductively guaranteed by the premises alone. The most prominent such inferences of course include cases of inductive generalization such as

$$\begin{array}{c} \textit{Many } F\textit{s are } G \\ \hline \therefore \textit{ All } F\textit{s are } G \end{array} \tag{1.3.1}$$

and cases of abduction (a.k.a. inference to the best explanation) such as

$$\begin{array}{c} E \\ \hline \textit{E is best explained by } T \\ \hline \therefore T \end{array} \tag{1.3.2}$$

as well as numerous variations on these themes. For such inference types the possible falsity

### 1.3. AMPLIATIVE ABSTRACTION AND EPISTEMIC DEBT

of the conclusion is not precluded (at least not on pain of inconsistency) by the truth of the premises. The premises instead are supposed to “support” the conclusion in virtue of some alternative method(s) or standard(s) of justification. So, if claims about phenomena indeed have content “additional” to the data records themselves as Woodward readily grants, it would seem that data to phenomena reasoning likewise counts as ampliative.

In order to get a sense of why claims about phenomena “go beyond” the content of the bare data records, let’s consider a few examples. For instance, as already discussed in the paradigm example of the melting lead phenomena, the quantity  $327.5^{\circ}C$  is recovered by means of the aggregation of numerous data records from individual samplings. But none of the individual data results may necessarily have the exact value of  $327.5^{\circ}C$ . Moreover, due to (almost inevitable) experimental noise, if such measurements are conducted with sufficient precision, we should expect that many of the data results differ (if only a little) from sample to sample. Nonetheless, an aggregation process such as calculating the arithmetic mean of the data results will recover the quantity  $327.5^{\circ}C$  as representative of *the* melting point of lead even though that specific number may not be equal to any one of the actual measurement results recorded in the data. By aggregating the data numbers to recover the claim about lead’s melting point, we have done more than simply restate the content of the data itself. To the contrary, it may be indicative of a claim about the phenomena that can easily differ from the particular data records. For a second example, consider a physicist conducting a double slit experiment with individually fired particles. The data recovered by such an experiment might be recorded as an array of numbers indicating the location where the individual particles irradiated the screen. What if the physicist next wants to make a claim about the *phenomena* concerning the probability distribution describing the chances of particles prepared in her experiment irradiating each region? In that case she will find a best fit curve

### 1.3. AMPLIATIVE ABSTRACTION AND EPISTEMIC DEBT

for the data recorded. However, there is a good deal of flexibility afforded to curve fitting analysis that is not determined by a set of data points alone. Nothing in the data alone determines which methods of interpolation one must use to best fit the plots: Should one use polynomials or special functions? Must we use only smooth functions, or differentiable functions, and if so to what order of differentiability? Shall we use piecewise interpolations? How many pieces, and with what kind of functions should the pieces be interpolated? As we will see in subsequent chapters when discussing the representation of thermal systems undergoing phase transitions, these questions are far from idle and are potentially influenced by background choices of theoretical context.

Though many sophisticated methods have been developed to guide data analysis within certain theoretical contexts, the question of how to best fit the data cannot be found in the experimental records alone. Moreover, even if this were not the case, a probability distribution for particles prepared in the experiment certainly involves claims about the likelihood of a particle hitting regions that were never struck in the course of the experiment, and hence not part of the recorded data. As a third example for the ampliative reasoning involved in data analysis, Woodward offers the analysis of fMRI data ([Woodward, 2011](#), p173). Raw fMRI measurement for each voxel can be especially noisy. To mollify such fluctuations in data records due to this noise, analysts make use of spatial smoothing procedures by averaging the value at each voxel with its neighbors. Again the analyst must go beyond the data records themselves in finding a way to describe the phenomena detected in a certain kind of smoothed form.

We might also consider a case recently emphasized by Sorin [Bangu \(2009\)](#) in discussing the difference between how thermodynamics and statistical mechanics are able to treat phase transitions. Phase transitions appear to manifest as certain “sudden changes” in the state

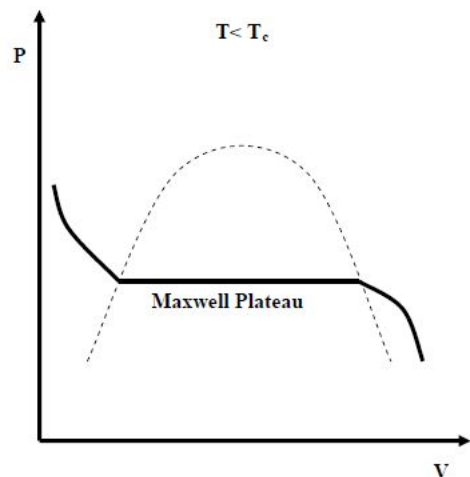


Figure 1.3.1: *Thermodynamic P-V curve representation of a system at fixed  $T < T_c$ .*

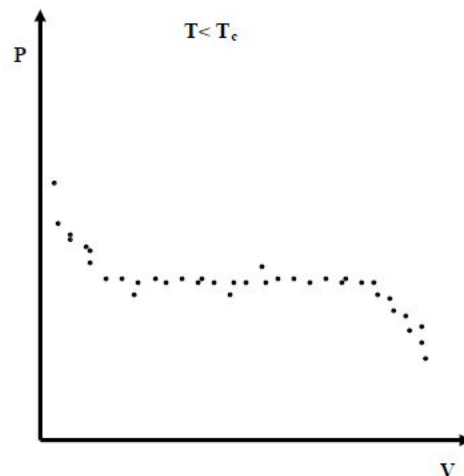


Figure 1.3.2: *Possible P-V data records of a thermal system held at fixed  $T < T_c$ .*

variables of the system (e.g. pressure and volume). In thermodynamics the ostensible suddenness of these changes is characterized by mathematical properties such as kinks and discontinuities in functions relating these variables. In figure 1.3.1 a thermodynamic depiction of the relation between pressure and volume of a thermal system held at constant temperature is given. As the curve crosses the region bounded by the dotted line a kink occurs indicating a (first-order) phase transition along the Maxwell Plateau.<sup>37</sup> A physicist, charged with the task of curve fitting data records as depicted in figure 1.3.2, might hence use such background theoretical presuppositions as a guide to fitting the curve, searching for least error solutions that fit the curve in three pieces (the middle of which is linear).<sup>38</sup> What is remarkable is that nothing about the actual data points can tell us anything about whether there is a kink or not in “the best fit” curve for the data. No matter how many measurements are taken, as long as we assume any measurement takes a non-zero amount

<sup>37</sup>The Maxwell Plateau indicates a state where the system is “infinitely” compressible (i.e.  $\frac{\partial P}{\partial V}|_T = 0$ ) occurring in a coexistence region between liquid and vapor states.

<sup>38</sup>This is essentially the kind of procedure used by thermal physicists (see (Malanowski, 1988, pp282-3)).



### 1.3. AMPLIATIVE ABSTRACTION AND EPISTEMIC DEBT

of time to conduct, there will only be a finite number of points. But a “kink” in the curve is a property of how the curve behaves as we continuously vary one of the state parameters. Hence, such data records alone cannot possibly indicate a kinky relationship between the pressure and volume of a thermal system. Further theoretical presupposition is required. As in the quantum example, not only is it the case that the analyst “fills in” places between the data points that were never actually measured, but she also may use significant theoretical presuppositions in guiding how to best fit the curve at all.<sup>39</sup>

Such curve fitting issues generalize well beyond our double slit and first-order phase transition examples. In any case where a scientist wishes to use a finite sampling of measurement results in order to detect properties of a scientific phenomena in the form a functional dependency, she will be faced with such data fitting challenges. But that means she must (a) go beyond the scope of the data itself to figure out what type of curve to fit the data with, and (b) she will have to fill in the untested “missing regions” between neighboring data entries. Hence, in moving from the raw data to more well behaved “smoothed out” functional descriptions of the phenomena, the analyst inevitably must go beyond the data itself. She must conduct ampliative reasoning in moving from the data records alone to claims about the phenomena.

Woodward notes that in the case of data to phenomena reasoning (as well as induction at large), further so-called “substantive empirical assumptions” must be adopted (either implicitly or explicitly) to license the drawing of the conclusion.<sup>40</sup> For instance, in the lead example, the choice to take the arithmetic mean of the data entries to determine the quantity

---

<sup>39</sup>In the case of the latter, the selection of certain *thermodynamic* presuppositions about phase transitions over those of statistical mechanics is stark, since it is impossible in the theoretical context of statistical mechanics to generate such kinks in representing the state variable relations of systems with finite degrees of freedom (see sections 2.2.3 and 2.3.2 in the next chapter).

<sup>40</sup>*Cf.* John Norton’s appeal to “material facts” to license ampliative inferences in his material theory of induction (Norton, 2003).

327.5°C as a description of the melting point of lead is supported not only by the data entries but by two further assumptions about the sources of error. Namely, the analyst assumes that the error effects are normally distributed (or at least symmetrical) and that their effects accumulate linearly.

Theoretical background assumptions about the type of physical system investigated are typically used to guide such ampliative inferences as well. In the quantum example, because it is common to make use of Hilbert space structures to describe the vector states of quantum systems, we might quickly judge that the proper curve to best fit our data with is the square of a Lebesgue square integrable function. That is to say, since the quantum mechanical theoretical context in which we describe wave functions makes use of normalized Hilbert space structures such as  $\mathcal{L}^2$ , with that background information we are easily guided into trying to best fit the data with that specific kind of function.<sup>41</sup> In this case, what happens is the ampliative reasoning from data to phenomenal description is informed by additional theoretical assumptions about the kind of system under examination.

Examples such as these are ubiquitous in science. Overwhelmingly, when a scientist makes the ampliative move of abstracting from raw data to a model that attributes the well specified interrelations of a phenomena, she appeals to a host of theoretical as well as empirical background assumptions. The point of this section and the chapter as a whole is not to impugn such practices. To the contrary, it does not appear that science could thrive at all without it. Instead our claim is that the process of abstraction both in extracting data and in moving from data to phenomena is guided by a number of non-apodictic influences independent of features directly manifested by the physical systems. In short, the thesis

---

<sup>41</sup>In quantum mechanics, (in part because of the Schrödinger equation) we typically go beyond just best fitting with normalized square integrable functions and whenever possible stick to even more well behaved functional descriptions such as smooth or at least doubly differentiable functions.

### 1.3. AMPLIATIVE ABSTRACTION AND EPISTEMIC DEBT

emphasized in this section is that *abstraction is ampliative*. Hence, in abstracting well defined phenomena potentially targeted by mathematical vehicles of representation, additional assumptions (theoretical and otherwise) are inevitably presupposed, incurring a kind of epistemic debt. Again, the claim is not a skeptical one: healthy scientific practice likely could not function at all without such presupposition. However, it is a mistake to trust our models without a recognition of the epistemic debt incurred through such ampliative abstraction procedures. In the next chapter, we shall focus on how such mistakes can occur when we take our mathematical vehicles to be representing more than the abstraction procedure licenses, and we will complete our development of a generalized account successful scientific representation designed to compensate for such potential errors.

## Chapter 2

# Scientific Representation and $\epsilon$ -Fidelity

In chapter 1, we began our investigation of how mathematics can be successfully applied to gain knowledge of the physical world. In particular, we began answering how scientists successfully use mathematical representations to gain knowledge of physical systems. In developing this answer we argued for the following: First, we saw that *scientific* representations used to successfully gain knowledge (i.e. faithful scientific representations) can face a “mysterious fidelity problem,” if the mathematical vehicle fails to capture any of the properties or interrelations exhibited by the targeted physical phenomena. We went on to argue in section 1.2 that while similarity matching is not necessary for representation, the mysterious fidelity problem may be resolved if the structure of the mathematical vehicle of representation does present a well defined abstraction of the relevant properties and interrelations of the physical system or systems. In section 1.3, we then argued that under such an account of successful scientific representation, the relevant interrelations abstracted from the physical systems targeted are not constituted by directly observed *data* records, but instead the detected *phenomena*. The chapter closed with the observation that, though the phenomenal patterns abstracted from experimental data clearly count as the relevant interrelations to be captured

by successful scientific representations, the epistemic procedure of extracting these patterns from the experimental data is ampliative. Abstracting the phenomenal patterns that can be well represented by the kind of mathematical representations useful to scientists is heavily dependent on background presupposition. In other words, the process of abstracting and then capturing phenomenal patterns useful for successful scientific inference incurs a kind of epistemic debt that we claimed should not be ignored when interpreting the significance of our mathematical representations.

In this chapter we will consider a further threat facing current matching solutions to the mysterious fidelity problem, arising from the widespread use of idealization in scientific representation. We will then complete our account of successful mathematical representation in science by developing a generalized<sup>1</sup> account of a class, called  *$\epsilon$ -faithful representations*, that may be legitimately applied to gain knowledge of physical targets. In particular, in this chapter we will show (1) that  $\epsilon$ -faithful representations are able to deal with the epistemological threat to soundness resulting from the use of both straightforward and pathological idealizations in science, (2) that  $\epsilon$ -faithful accounts of representation can be no less confirmed by examples of successful mathematical application than traditional matching accounts, and (3) that much of the “epistemic debt” that was highlighted in section 1.3 as accruing during the abstraction process can be eliminated when the success of our representations are understood under the  $\epsilon$ -fidelity account.

---

<sup>1</sup>Note, it is consistent with this thesis both that alternative solutions to the mysterious fidelity problem exist and that accounts of representation used under such solutions need not satisfy  $\epsilon$ -fidelity. That is to say, we are not claiming that the account of  $\epsilon$ -faithful scientific representation argued for in this chapter is the *only* possible resolution of the mysterious fidelity problem posed in section 1.2.1. (Though it is difficult to imagine how such alternative accounts might function.) Rather, we argue that accounts of representation that take advantage of some sort of matching relationship to avoid this problem as developed in section 1.2 must allow for the broader conception of “matching” developed below in section 2.1.2. Moreover, the  $\epsilon$ -fidelity solution completed in this chapter offers a general method of broadening this conception without running into the sorts of errors that will be introduced in section 2.1.1 or highlighted in more complex cases of universality discussed in chapter 4.

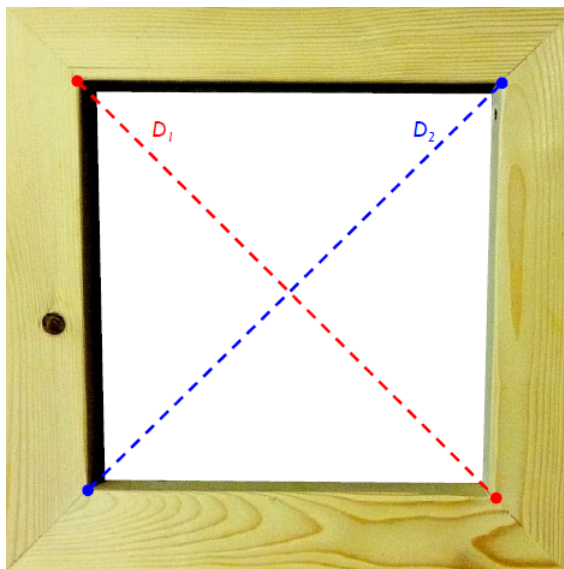


Figure 2.1.1: *Physical wooden frame with interior diagonals  $D_1$  and  $D_2$ .*

## 2.1 Idealization and Plato's Problem

### 2.1.1 Framing the Problem

In this section we will consider a potential problem associated with misinterpreting the significance of abstracted models. Specifically, we will be concerned with understanding the role of models in cases where **(a)** the mathematical representation can be used to gain *knowledge* about target physical systems, but **(b)** the properties of the abstracted model must either fail to preserve or directly misrepresent certain *relevant* properties or relations present in the actual physical systems in order to gain such knowledge. We begin with a simple example:

A carpenter is framing a square shaped window, and wants to make sure that the interior corners are joined at right angles. She proceeds to measure the diagonals of the interior as in figure 2.1.1, to recover the lengths of  $D_1$  and  $D_2$ . Observing that the measured values are

## 2.1. IDEALIZATION AND PLATO'S PROBLEM

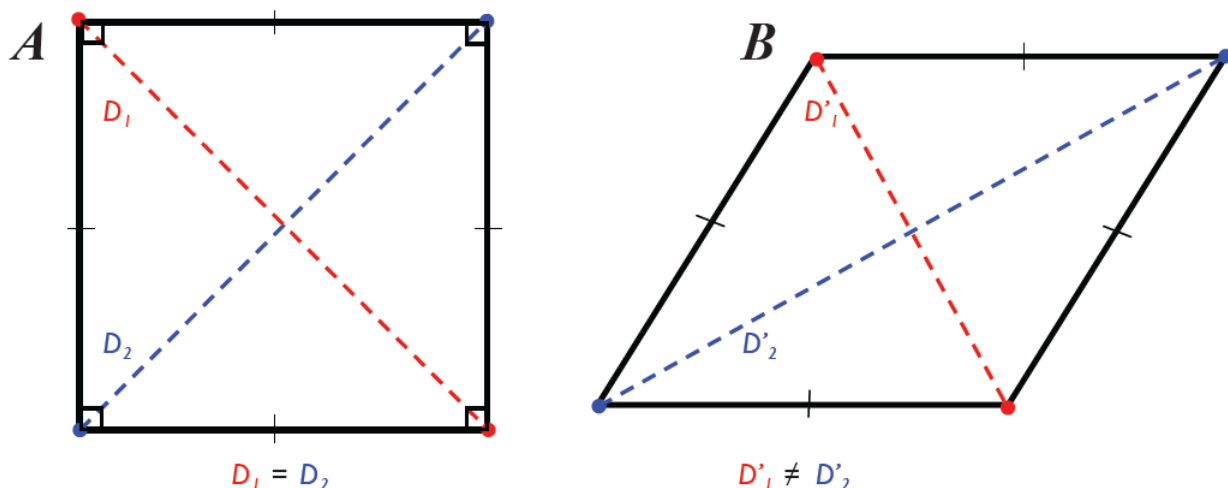


Figure 2.1.2: *Quadrilaterals A and B with congruent side lengths.*

equivalent, she now knows that the interior corners have been joined at right angles.<sup>2</sup> How does she know this?

An apparent justification of this knowledge can be given with a “matching account” of mathematical representation of the sort discussed in section 1.2. She first models the interior edges of the frame as a geometric figure in a two dimensional Euclidean plane. Since she cut all four frame pieces to the same dimensions, she matches this (relevant) property by requiring that that all four edges of her abstract geometric model have congruent side lengths (i.e. the model is a rhombus). She can now make use of the following elementary theorem of Euclidean geometry:

**Theorem 2.1. (*Congruent Diagonals Theorem*)** *Let  $\mathbf{X}$  be any rhombus, then the diagonals of  $\mathbf{X}$  are congruent if and only if all of  $\mathbf{X}$ ’s interior angles are right angles.*

The congruent diagonals theorem allows our carpenter to distinguish whether her geometric model representing the wooden frame is like object **A** or object **B** in figure 2.1.2.

---

<sup>2</sup>This “diagonals measuring” technique is an actual practice well recognized in cabinetry and framing as more reliable (and convenient) than direct angle measurement.

## 2.1. IDEALIZATION AND PLATO’S PROBLEM

Object **A** has congruent diagonals so according to the theorem, it must have only right angles. It is a square. Object **B** on the other hand has diagonals that are not congruent. Hence, its angles cannot be right, and it is a mere rhombus. The reasoning behind our carpenter’s conclusion that the physical frame has right angles now works as follows: Because the wooden frame has congruent diagonals, she matches this (relevant) property in the abstract geometric model representing the frame, requiring it to also have congruent diagonals. So by the congruent diagonals theorem, she can make the mathematical deduction that the geometric model must also have right angles. Matching this (relevant) right angles property of her geometric model with the angle measure properties of the wooden frame it represents, she concludes that the actual wooden frame also has interior angles that are right.

This justification, much like the Königsberg bridge example of section 1.2.3 above has a lot of intuitive appeal. The justification appears only to rely on matching unequivocally “relevant” properties of the physical wooden frame with the abstract geometric model. The reasoning schema is straightforward: First, develop a mathematical model that matches the relevant properties exhibited by the physical target. Then, use these matched properties in the mathematical model to deduce further properties that the mathematical model must exhibit. Last, determine which physical properties would be needed to match the newly deduced mathematical properties, and conclude that the physical system must also have these properties. The reasoning is diagrammed in figure 2.1.3. This “C-shaped” inference schema is typical of many matching accounts found in the literature.<sup>3</sup>

The problem with this inference schema is that the “matching inferences” (diagrammed by the horizontal arrows) frequently fail. That is to say, *the relevant properties of the*

---

<sup>3</sup>See e.g. (Bueno & Colyvan, 2011, §4) or (Hughes, 1997, §2) and their respective IDI and DDI schemas.



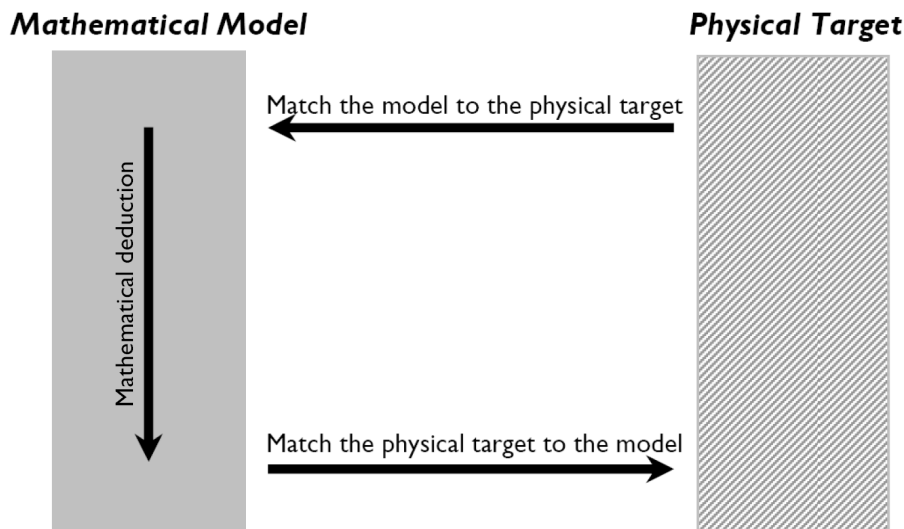


Figure 2.1.3: “C-shaped” inference schema found in matching accounts of scientific representation.

mathematical model do not match the relevant properties of the concrete physical system. The properties don’t actually match. This mismatching problem is evident even in our simple carpentry example. In order to draw the conclusion of the congruent diagonals theorem the mathematical figure must not only have congruent diagonals, it must be a rhombus. Rhombi are necessarily bounded by four edges that are all perfectly straight and perfectly congruent to one another.<sup>4</sup> So, in order for the above elaborated mathematical deduction to be sound, we invoked a mathematical model with perfectly straight and congruent edges. On the other hand, these very properties of the mathematical model were supposed to match up with the corresponding physical properties abstracted from the actual wooden frame.

The problem is that in order to make a sound mathematical deduction scientists frequently appear to match mathematical models meeting very specific constraints with physical systems that fail to exhibit such properties. The “messy” properties of the physical

---

<sup>4</sup>Theorem 2.1 of course generalizes for non-rhombus parallelograms, which need not have four congruent edges. However, their opposing edges must still be “perfectly” congruent.

## 2.1. IDEALIZATION AND PLATO’S PROBLEM

system are matched with the more ideal properties imposed on the mathematical model. In deference to his classic attention to the mismatch between mathematical structures and the world of experience, we will refer to this epistemological challenge to soundness resulting from relevant properties of a represented physical system not meeting the idealized constraints of a mathematical model as *Plato’s problem*.

The failure of physical systems to meet the precisely defined constraints imposed on their mathematical models is almost ubiquitous in science. Of course, if a property (or relation) is *irrelevant* to deductions made using the representation, it is unproblematic when such properties of the physical system fail to be recovered in the model.<sup>5</sup> Mismatching becomes a problem, however, when the very properties that are used in the mathematical deduction step of the “C-shaped” inference pattern are not or cannot be matched by the physical system. In such cases, matching accounts are faced with a dilemma: either the mathematical deduction is unsound, or the matching inferences cannot apply to the actual physical targets. Plato’s problem poses an epistemological challenge that must be met by any account of faithful (knowledge enabling) scientific representation that relies on some sort of “matching” solution to the mysterious fidelity problem. Unfortunately, accounts in the literature appealing to matching solutions are uniformly ill equipped to deal with this challenge.<sup>6</sup>

---

<sup>5</sup>E.g., in the Königsberg bridge example of section 1.2.3, it was irrelevant how wide each of the seven bridges were, so it did not matter that such information was not captured in the abstract bridge structure depicted in figure 1.2.2.

<sup>6</sup>Pincock observes that what he calls “matching models” (models with properties that do match up with those of an individual physical target) may be related to the idealized models (called “equation models”) used for drawing mathematical inferences, by means of an “acceptable mathematical transformation,” where “[a] mathematical transformation will be acceptable when it is consistent with the goals of the scientists in terms of scale and accuracy” (Pincock, 2007b, p963). However, he does not offer any method of evaluating the epistemological legitimacy of such transformations or an account of why a transformation may (or may not) be trusted as staying consistent with such desiderata. Bueno & Colyvan (2011) develop an elaborate account of what they describe as “partial” isomorphisms which preserve some but not all of the relevant constraints potentially matched with the physical target, but then fail to offer any account of why the mismatched relevant properties may be legitimately used at the stage of mathematical deduction to gain knowledge of the actual target. (Cf. further critiques found in (Batterman, 2010).)

## 2.1. IDEALIZATION AND PLATO’S PROBLEM

One reason for the dearth of acceptable solutions has to do with the concept of ‘idealization.’<sup>7</sup> There is a strong temptation, especially in cases such as our above simple carpentry example, to dismiss the dilemma and ignore “minor mismatching” as cases of simple idealization. While in some cases, as the below solution of section 2.3 will elaborate, idealization can be invoked in such an exculpatory fashion, Plato’s problem reveals a very direct epistemological threat. As we shall see in section 2.3 this threat becomes especially vivid in examples that we will refer to as *pathological idealizations* in which a representation uses constraints on a particular mathematical model to gain knowledge of a physical system that either **(I)** cannot possibly be realized by any physical system or **(II)** render the mathematical model incompatible with the defining constraints of the theory under which the model is invoked. Our solution to Plato’s problem presented below in section 2.2.1 is aimed not only at resolving the often overlooked epistemological deficiency of matching accounts resulting from minor idealizations (as in our carpentry example) but more importantly those deficiencies involved in the abundant use of such pathological idealizations in the mathematical sciences.

### 2.1.2 Scope and Structure

In order to develop our solution to Plato’s problem presented below, in this section we will take a look at precisely what kind of properties may ostensibly be “matched” in general mathematical representations. An apparent candidate would be something like “the relevant properties of the target being represented,” where relevance is cashed out in terms of the intrinsic features of the system of interest to the representing agent(s) (e.g. the physicist).

Matching the intrinsic properties of the target to the internal relations of a mathematical

---

<sup>7</sup>The concept ‘idealization’ like the concept ‘representation’ suffers from a significant lack of univocal usage. In the present context we invoke the term ‘idealization’ broadly to include alternative concepts such as ‘approximation’ that are also occasionally (though again not univocally) employed to indicate a kind of “legitimate” mismatching of mathematical models with physical targets in scientific representation.

## 2.1. IDEALIZATION AND PLATO'S PROBLEM

model can be useful when the model itself has rich internal structure. However, in some cases, the internal relations of the target and the corresponding model are not the only features of interest. That is to say, the “internal structure” of an object is not the only structure to be tracked.

To bring out this distinction, let us consider Pincock’s example of counting apples on a table. He describes in this example the process of assigning natural numbers (in order) to the respective apples. When there are no further apples to which we can assign a further number, the apples are counted and the last assigned number tells us how many apples are on the table:

What is going on here? A natural thing to say is that there is a mapping of a specific kind from the apples on the table to an initial segment of the natural numbers. This mapping is called an isomorphism. Briefly, an isomorphism is a mapping that preserves cardinality and structure. Now, when I count the apples I am determining that there is an isomorphism from the apples to the natural numbers starting with 1. We can capture, then, the kind of external relation that is required by talking of mappings and their properties. Here we have a statement of the form ‘There are  $n$   $F$ s’ coming out true just in case there is an isomorphism from the  $F$ s to an initial segment of the natural numbers ending with  $n$ . (Pincock, 2004, pp145-6)

Pincock describes his appeal to isomorphic matching in mathematical representation as the “structuralist approach.”<sup>8</sup> The relevant “structure” of *internal* relations of the apple system (e.g. how one apple relates to an other) bear an *external* relationship to some mathematical system. The external relationship is given by the existence of a mapping relationship, an isomorphism from the apples on the table onto a particular (ordered) initial subset of the natural numbers.

As discussed in section 1.2.2 above, establishing an isomorphism from physical “apples

---

<sup>8</sup>Structuralism in mathematical representation is not to be confused with structural realism. The latter pertains to a kind of ontological position one might take with regard to mathematics, whereas the former pertains to applying mathematics and does not take a position on the ontological status of the mathematics being applied.

## 2.1. IDEALIZATION AND PLATO'S PROBLEM

on the table” to natural number sub-sequences constituted a kind of category error. The specific apples on the table and their “structure” is not (yet) well defined. It was argued there that, instead of mapping concrete physical systems or components of the physical systems, we must *abstract* the relevant and stable properties from the physical system(s) or system type(s) in question. In this counting example, it appears that the relevant stable structure preserved by a Pincock isomorphism ( $\varphi_n$ ) is constituted by the non-identity relations that obtain between the various apples on the table. That is to say, a non-identity relation obtains between any two distinct apples that are on the table. Other properties or relations that hold of apples on the table (such as which one is the largest apple, which ones are ripe, etc.) are ignored as irrelevant to the structure preserved by  $\varphi_n$ . So, an abstraction of a given batch of apples may be constituted by a set of non-identical apple tokens, one for each distinct apple on the table. Such an abstract apple-batch set represents the domain of a particular Pincock isomorphism  $\varphi_n$ .<sup>9</sup> This abstracted “apple-batch structure” is matched through a  $\varphi_n$  with a well defined (and structured) initial subset of the natural numbers.

Pincock is correct that such a bijective (i.e. one-to-one and onto) mapping from the batch of distinct apples (or at least an abstraction of their mutual distinction) to initial sub-sequences of the natural numbers preserves the cardinality of the domain once we move to the range. That is to say, such a bijection exists if and only if the domain (the abstraction of the apples) has the same cardinality as the range (members of an initial subset of the natural numbers). Hence, since the largest member of such initial sequences specifies the cardinality of the “segments,” this largest number must also specify the cardinality abstracted from the apple system in question. If all we are concerned with is assigning different numbers to

---

<sup>9</sup>Note, Pincock’s construction does not actually specify a single isomorphism  $\varphi$  in the above long quote but a multitude of isomorphisms  $\varphi_1, \varphi_2, \dots$  where the range of each distinct mapping  $\varphi_n$  is the initial sequence of natural numbers ending with  $n$  and the domains of the respective isomorphisms are constituted by batches of  $n$  distinct apples (or as we have argued, abstractions of these respective batches).

## 2.1. IDEALIZATION AND PLATO’S PROBLEM

different amounts of apples, then a mere bijection between abstractions of the apple systems and initial sequences of different cardinality does the trick. However, when we say, for example, that “there are 8 apples on this table” or “there are 4 apples on that table,” such an application of numerals typically is intended to entail more than the mere fact that the number, 8, assigned to the first batch of apples is not identical to the number, 4, assigned to the latter batch. There is usually more “structure” implied when we count batches of apples.

Each of the  $\varphi_n$  bijections establish that there are “ $n$   $F$ s on the table” on what van Fraassen refers to as a merely *nominal scale*: It assigns numerical labels to different sets of  $F$ s “without implying any algebraic structure” (van Fraassen, 2008, p116). An example of such a (merely) nominal assignment is the numbering of players on a sports team. A baseball player with the number 8 is not so labeled because he is “twice as good” in some respect when compared to a player with the number 4, nor does it suggest that the player with the lower number is better (or worse) than a player with a higher number. The numbers are simply assigned to provide a “nominal” distinction.

In contrast, when we make such claims about the number of apples on tables we typically are making claims that do imply further facts: A table with 8 apples has *more* apples than the table with 4 apples; in fact, the former has *twice as many* apples as the latter; both sets of apples can be divided into two equal groups whereas a table with 9 apples cannot be, etc.

It is arguable that the structure preserved in Pincock’s original counting example (contrary to our above reconstruction) is not merely nominal, but also establishes an *ordering* among various batches of apples. After all, there do exist well defined orderings on the finite cardinal numbers. What kind of relationship would such an ordering mean with respect to sets of apples? It would be a set of relations that exist not between different apples on a ta-

## 2.1. IDEALIZATION AND PLATO'S PROBLEM

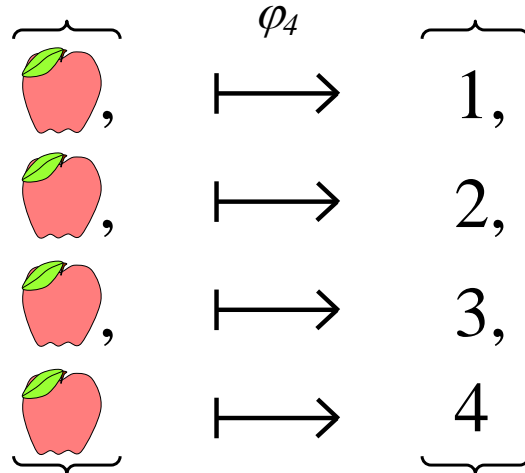


Figure 2.1.4: The mapping  $\varphi_4$  from an abstract batch of four apples onto the set  $\{1, 2, 3, 4\}$ .

ble, but between *different potential batches of apples*.<sup>10</sup> For instance, it would establish that our table with 4 apples bears a certain relationship (a fewer than relationship) to the table with 8 apples. But recall, the “isomorphism” that Pincock identifies is *not* a mapping from the set of sets of potential apple-batches (or their abstractions) to a structured set of initial subsets of the natural numbers as in figure 2.1.5. Rather, it was a mapping from *apples in a batch* (or better yet their abstraction) to *numbers in an initial segment* as in figure 2.1.4. The members in set  $A$  of figure 2.1.5 are abstractions of entire batches of apples, and the members in set  $B$  are initial subsets of the natural numbers. The Pincock isomorphisms  $\varphi_n$  do not preserve the structure that exists between members of set  $A$  by mapping those members (in an ordered preserving way) to the members of the set  $B$ . The scope of the structure

<sup>10</sup>The concept ‘potential’ used here is quite broad. It includes actual and physically possible batches of apples. However, it may also include apple systems that are, in some sense, not physically possible. For example, if there is a finite amount of matter-energy in the universe, then according to any modern physical theory with a law of conservation of matter-energy, there is an upper limit  $N$  on the number of apples that could exist in our universe. Despite this fact, in some cases it may be epistemologically legitimate to draw ampliative inferences (based on relationships between apple systems with less than  $N$  apples) that, for example, a batch of 4 apples bears other relationships to systems with more than  $N$  apples, even if the latter system is (in some sense) physically impossible. Ultimately, an answer to what counts as a “potential” or “possible” system will be dependent on the ampliative inferences involved in abstracting this kind of extrinsic structure.

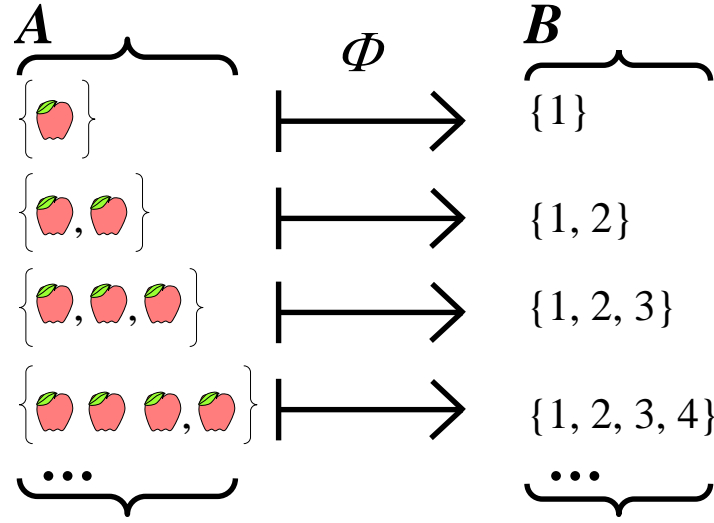


Figure 2.1.5: A mapping  $\Phi$  from the set  $A$  of abstract apple-batch sets onto the set  $B$  of initial subsets of the natural numbers.

preserved by Pincock isomorphisms is too narrow. The bijection  $\varphi_4$  maps elements from a single member of  $A$  (i.e. the abstraction of a table with 4 apples) to a particular subset of the natural numbers (i.e. the ordered set  $\{1, 2, 3, 4\}$ ). But such an isomorphism can only preserve the *intrinsic* relations abstracted from the apples on the table, the *intrinsic structure*. If we want to preserve (and then make use of) facts about how the number 4 relates to the number 8, we must preserve *extrinsic structure* too. The structured domain of the mapping must include not just relations between the 4 apples on the table but, for example, how those apples relate to a potential table with 8 apples. If we want to be able to imply something about the “order relations” that may also obtain between a batch of 4 apples and a batch of 8 apples, *we must expand the scope of the domain and structure preserved by the mapping*. The internal structure abstracted from a particular batch of apples is not the only thing that can be preserved. If we want to help ourselves to the rich algebraic and ordering relations that can be imposed on the natural numbers (or isomorphically equivalent sets like  $B$ ),<sup>11</sup> the “mapping” appealed to must preserve not merely the (intrinsic) relations that exist

<sup>11</sup>Note, the set  $B$  is not technically the set of natural numbers  $\mathbb{N}$ . However, it is trivial to define a desired



## 2.1. IDEALIZATION AND PLATO'S PROBLEM

among the individual apples on the table, but the *extrinsic* relations that the batch on the table has with respect to other potential apple batches.

By preserving an abstraction of these external relationships that exist between different potential apple-batches, like a batch of 4 apples and a batch of 8, a mapping like  $\Phi$  depicted in figure 2.1.5 preserves more than the fact that the first batch has a “different amount” than the second. It allows us to take advantage of a number of the other interrelations in the algebraic and ordering structure of the natural numbers. This means that a “large scope” structure preserving map like  $\Phi$ , which takes into account the external relationships to other apple batches, allows us to gain knowledge that would not be justified by the Pincock isomorphisms. Unlike before, when this larger scope structure is preserved, claims like “there are 4 apples on the table” can now be used to conclude that “if we eat half the apples there will be 2 left,” because, in particular, the relevant extrinsic apple-batch relations were preserved in the algebraic relationship  $4 - \frac{4}{2} = 2$ .

The point generalizes: representation with well defined, structured mathematical systems allows us to formalize physical relations as a rich network of relationships that exist among the members of the mathematical systems. But if the knowledge we wish to gain from these mathematical representations is going to hinge on this rich structure, then an account of mathematical representation that appeals to “relationship preserving” matching techniques like (isomorphic or homomorphic mappings) can error if the scope of the representational target’s preserved structure is too narrow. In the Pincock example, we may want to count apples on the table to draw a number of conclusions, but many (perhaps most) of the conclusions we would like to draw when we note that there are “4 apples on the table” do

---

algebraic and ordering structure of the natural numbers on a set like  $B$  that would enable the existence of an isomorphism between the two. We hence proceed under the assumption that  $B$  is so structured.

## 2.1. IDEALIZATION AND PLATO'S PROBLEM

not require merely matching “the number 4” with an abstraction of the apples on the table. They require matching a whole host of numbers (e.g. all the natural numbers) with a host of potential apple-batches, one of which happens to be instantiated by the actual batch on our table. It is only by enlarging the scope of the representation’s target to include not only the apples on the table but their external relationships to other “apple systems” that such a (matching dependent) representation can hope to accomplish the epistemic goals we expect. As we shall see, such scope enlargement, especially when it comes to preserving extrinsic “similarity” relationships in the form, for example, of topological and metric relations in the mathematical structure of a vehicle of representation, will be vital to resolving Plato’s problem in order to legitimize the use of idealizations.

### 2.1.3 Close Enough for Carpentry

We saw in section 2.1.1 that Plato’s problem arises for strict matching accounts of mathematical representation whenever the mathematical models used to draw conclusions about the physical system fail to precisely match features of the physical system relevant to the mathematical deduction. The problem, exemplified by our carpentry example, was that some of the mathematical model’s conditions that were essential to the deduction are not met by the actual wooden frame that the model was supposed to represent. Such a matching account faces the following dilemma: either the deduction is not sound (the mathematical conditions are not actually satisfied by our model) or the model does not apply to our physical target (it cannot be used to gain knowledge about the system in which we are interested). Since such mismatching between model and physical system is nearly ubiquitous in the mathematical sciences (more the rule than the exception), this dilemma poses a significant epistemological deficiency for such matching based accounts of mathematical representation or application.

## 2.1. IDEALIZATION AND PLATO'S PROBLEM

Though the significance of this dilemma should not be overlooked, there is a compelling temptation to insist that Plato's problem is only a problem if we "take our matching expectations too seriously." Recall, an alleged mismatching observed in section 2.1.1 was that the edges of the wooden frame were not "perfectly congruent" and hence could not match the perfect congruency properties of a geometric figure like a rhombus. In response to this observation it is difficult to deny that despite this lack of "perfect matching" the wooden frame can be made to come *close* to being congruent. This temptation to rely on closeness is the key to resolving Plato's problem. The challenge facing a matching account of representation is to build a way of keeping track of relevant closeness into our matching inferences and, when appropriate, to avoid drawing conclusions about a physical target that can only succeed in the case of perfect matches.

The primary step in resolving Plato's problem is finding a way to keep track of close but imperfect matches between the models used for a mathematical deduction and the actual physical systems about which we wish to gain knowledge. In section 2.1.2 we saw that when the matching relationship is prescribed too narrowly (e.g. between a *single* model and an abstraction of a individual physical system), we lose the ability to draw all of the mathematical deductions that might be made to gain knowledge about the physical target. Instead, it was argued that in some cases we should expand the scope of such matching relationships, matching not just individual physical system abstractions with individual mathematical objects or elements (i.e. individual models), but abstractions of the physical system and all of the extrinsic relationships it potentially bears to other relevant physical systems with a *set* or *space of models*. If we move to this kind of larger scope matching then we have the ability to gain knowledge entailed not just by a matching of internal relations that are preserved when we move from individual system to individual model.

## 2.1. IDEALIZATION AND PLATO'S PROBLEM

Our solution to Plato's problem relies in particular on keeping track of external "closeness" or "similarity" relationships that exist, on the one hand, between actual physical systems and physically idealized versions of those systems, and, on the other hand, between perfectly matched mathematical models and the kind of idealized mathematical models frequently employed in mathematical applications.

To see how this works, let us return to our carpentry example. According to our story, the carpenter cut each of the wood pieces by the same method, and so after joining the interior edges of the frame our carpenter concluded that they should be of "equivalent length." However, even with the best carpentry techniques, it is not certain that the lengths of the interior edges of the wooden frame must be *exactly* the same length. Hence, it was suggested that modeling the frame with a rhombus with four perfectly congruent sides may have involved a mismatching. Instead of modeling the frame with a figure that has four perfectly congruent sides of length equal to exactly  $S$ , the frame might be less mismatched if it were modeled by some quadrilateral with edges of side lengths that are "close to  $S$ ."

For every fixed  $\delta \in [0, 1)$  and  $S \in \mathbb{R}$ , let us consider the set  $\mathcal{V}_{\delta, S}$  of quadrilaterals with congruent diagonals of measure  $D_1 = D_2 = S\sqrt{2}$ . The four side lengths of a member in  $\mathcal{V}_{\delta, S}$  are determined by multiplying the "original side length"  $S$  by a factor  $(1 + \delta_i)$  where  $\delta_i \in (-\delta, \delta)$  for  $i = 0, 1, 2, 3$  (see figure 2.1.6). For sufficiently small  $\delta \geq 0$ , the members of a set  $\mathcal{V}_{\delta, S}$  will all have side lengths that are "close" to congruent (differing from one another in measure by at most  $2\delta S$ ) and close to the idealized case in which  $\delta_i = 0$  for all four sides. For fixed  $S$ , each choice of  $\delta$  determines a well defined "closeness neighborhood"  $\mathcal{V}_{\delta, S}$  of models that are " $\delta$ -close" to the idealized model with congruent diagonals and fixed side lengths of exactly  $S$ , and the smaller  $\delta$  gets, the closer the members in the neighborhood are

## 2.1. IDEALIZATION AND PLATO'S PROBLEM

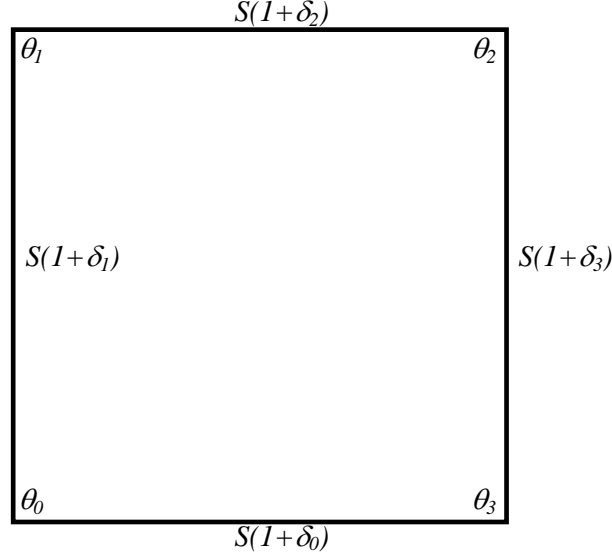


Figure 2.1.6: A member of the set  $\mathcal{V}_{\delta,S}$  with congruent diagonals of measure  $S\sqrt{2}$  and  $\delta_i \in (-\delta, \delta)$  for  $i = 0, 1, 2, 3$ .

guaranteed to having the perfect congruency property found in the idealized model.

What happens if our representation keeps track of these “extrinsic” relationships between different models in such spaces of quadrilaterals? In particular, we would like to know if we can gain any knowledge about whether members of a sufficiently narrow neighborhood  $\mathcal{V}_{\delta,S}$  must have right interior angles. The answer is “no but we can make them come close.” More precisely, let  $\epsilon_i := |\theta_i - \frac{\pi}{2}|$  for each of the interior angles  $\theta_i$  of a given “ $\delta$ -close” quadrilateral in  $\mathcal{V}_{\delta,S}$ , then it follows from the law of cosines<sup>12</sup> that  $\epsilon_i \leq \frac{\pi\delta(\delta+2)}{2(1-\delta)^2}$  for all four angles. So, if we want a model with angles that differ from  $\frac{\pi}{2}$  by no more than  $\epsilon$ , then there is a neighborhood

---

<sup>12</sup>For  $i \equiv 0, 1, 2, 3 \pmod{4}$  and  $\delta_i \in (-\delta, \delta)$ , then from the law of cosines we get:

$$\cos(\theta_i) = f(\delta_i, \delta_{i-1}) := \frac{\delta_i^2 + \delta_{i-1}^2 + 2(\delta_i + \delta_{i-1})}{2(1 + \delta_i)(1 + \delta_{i-1})}.$$

So if  $|\delta| < 1$ , we can deduce the following inequality:

$$\begin{aligned} \epsilon_i &:= \left| \theta_i - \frac{\pi}{2} \right| = |\arcsin(f(\delta_i, \delta_{i-1}))| \leq \left| \frac{\pi}{2} f(\delta_i, \delta_{i-1}) \right| \\ &\leq \frac{\pi}{4} \frac{|\delta_i^2| + |\delta_{i-1}^2| + 2|\delta_i| + 2|\delta_{i-1}|}{(1 - |\delta_i|)(1 - |\delta_{i-1}|)} \leq \frac{\pi\delta(\delta + 2)}{2(1 - \delta)^2}. \end{aligned}$$

## 2.2. SOLVING PLATO’S PROBLEM

of models  $\delta$ -close to the square with side-length  $S$  (i.e. the set of models  $\mathcal{V}_{\delta,S}$ ), whose angles are guaranteed to be that close or better.

This  $\epsilon$ - $\delta$ -relationship gives rigorous support to the above mentioned temptation to respond to Plato’s problem by pointing out that though the model does not *precisely* match the wooden frame, it is “close enough.” What the above analysis demonstrates (in a sense) is how close “close enough” actually should be. If we want our angles to be less than a certain number of radians from  $\frac{\pi}{2}$  all we have to do is make sure that the variation from congruency of the edge lengths is no worse than a certain  $\delta$ . As long as our wooden frame is well matched with *some* member of  $\mathcal{V}_{\delta,S}$  the above analysis ensures that the angles of our wooden frame must be  $\epsilon = \frac{\pi\delta(\delta+2)}{2(1-\delta)^2}$  close to being right (or better).<sup>13</sup> Hence, though we *cannot* use our geometric representation to guarantee that the interior edges of our wooden frame have been joined at *exactly* right angles, by taking advantage of the  $\epsilon$ - $\delta$ -relationship, we can still gain knowledge related to this claim. By using a representation that keeps track of the extrinsic  $\delta$ -neighborhood structure we *can* gain knowledge that the angles are  $\epsilon$ -close to being right.

## 2.2 Solving Plato’s Problem

### 2.2.1 $\epsilon$ -Faithful Representation

The idealized rhombus representation from section 2.1.1 matched in isolation runs afoul of Plato’s problem. The soundness dilemma generated by this problem prohibits gaining the

---

<sup>13</sup>The reader may be bothered by the constraint imposed on membership into  $\mathcal{V}_{\delta,S}$  that  $D_1 = D_2 = S\sqrt{2}$ . After all it is as suspect to assume that the diagonals are congruent as it is to claim that the sides are congruent. For the purpose of simplicity, we left out this constraint, but it is not difficult to verify by a calculation similar to the one used in note 12 that if we instead considered sets of the form  $\widehat{\mathcal{V}}_{\delta,S}$  such that not only were the side lengths  $S_i = S(1 + \delta_i)$  perturbed by suitably small  $\delta_i \in (-\delta, \delta)$  for  $i = 0, 1, 2, 3$ , but also the diagonal lengths  $D_j = S\sqrt{2}(1 + \widehat{\delta}_j)$  where  $\widehat{\delta}_j \in (-\delta, \delta)$  for  $j = 1, 2$  then the upper bound on each error term  $\epsilon_i$  increases by at most 50%. That is to say, the interior angles of each member in such a  $\widehat{\mathcal{V}}_{\delta,S}$  differ from  $\frac{\pi}{2}$  by no more than  $\frac{3}{2} \cdot \frac{\pi\delta(\delta+2)}{2(1-\delta)^2}$ .

## 2.2. SOLVING PLATO’S PROBLEM

knowledge that the wooden frame has right angles and is hence *not* epistemically faithful. However, we saw in section 2.1.3 that if we use a larger scope representation, matching our actual wooden frame and its extrinsic relationships to other potential physical frames with the  $\delta$ -neighborhood structure encoded through membership in the various  $\mathcal{V}_{\delta,S}$  sets, we could deduce the knowledge that our wooden frame has right angles *up to an  $\epsilon$ -error*.

Let us call representations that enable knowledge of a target up to a specified margin of error (epistemically)  *$\epsilon$ -faithful representations*. As in the carpentry example,  $\epsilon$ -faithful scientific representation occurs whenever it is possible to deduce the knowledge that a physical target is “ $\epsilon$ -close” to having certain well defined properties. To be precise, let  $S$  be some structured set of mathematical models such that the structure on  $S$  and the properties of the elements of  $S$  match certain relevant extrinsic and intrinsic relations that might be abstracted from a potential physical target system type. For example, in the carpentry case,  $S$  might be the set of all Euclidean quadrilaterals with congruent diagonals.<sup>14</sup> Let a structured mathematical space  $P$  be referred to as a *property space* of  $S$  if there exists a *P-property mapping*  $\varphi : S \rightarrow P$  such that  $\varphi$  is a homomorphism with respect to a given subset of the total set of relations (i.e. the structure) defined on  $S$ . In other words, this subset of relations is preserved by  $\varphi$  in the structure defined on  $P$ .<sup>15</sup>

Observe, property mappings are not necessarily surjective. That is to say, there may be elements in  $P$  that do not exist in  $\varphi[S]$ .<sup>16</sup> If part of the structure defined on a given property

---

<sup>14</sup>As discussed in note 13,  $S$  could have been broadened so as to weaken the congruent diagonals constraint.

<sup>15</sup>Note, unlike the “homomorphic mappings” from physical systems to mathematical systems, property mappings do not run into van Fraassen’s category error challenge of section 1.2.2 because  $S$  is a structured set of mathematical elements, already abstracted from some physical target or targets. Hence, both the domain  $S$  and the range  $P$  are well defined structured sets.

<sup>16</sup>For example, let  $S$  be the set of all finite subsets of (ruler and compass) *constructable* vertices in  $\mathbb{R}^2$  and let  $P$  be the space of real numbers  $\mathbb{R}$  with all of the algebraic relations and structure entailed by those relations. We can define one possible  $P$ -property mapping  $\varphi$  such that for a given  $s \in S$ ,  $\varphi(s)$  is the distance between the farthest two vertices of  $s$ . Since all vertices in  $s$  must be constructible points in  $\mathbb{R}^2$ , it is provable that the image  $\varphi[S]$  consists of only elements in the field of constructable real numbers. Hence, the *non-constructable* number  $\sqrt[3]{2} \notin \varphi[S]$  even though it is clearly a real number and hence in  $P$ .

## 2.2. SOLVING PLATO'S PROBLEM

space  $P$  includes a topological structure  $\tau_P$  (or, of course, any richer structure entailing the existence of a topological structure, such as a metric structure), then we can speak rigorously of closeness relationships of members of  $P$  in terms of these topological relations.<sup>17</sup> Hence, if  $P$  is a property space of a space of models  $S$  with topology  $\tau_P$ , given some “property”  $p \in P$ , for any model  $s \in S$  we may rigorously state how close  $s$  is to having property  $p$  by looking at which of the *neighborhoods* of  $p$  (i.e. the topological elements  $\epsilon(p) \in \tau_P$  such that  $p \in \epsilon(p)$ ) contain the  $\varphi$ -image of  $s$  in the property space (i.e. which  $\epsilon(p)$  contain the element  $\varphi(s) \in P$ ). For every neighborhood  $\epsilon(p)$  such that  $\varphi(s) \in \epsilon(p)$  we may say that  $s$  is  $\epsilon$ -close to having property  $p$  (with respect to the  $P$ -property mapping  $\varphi$  and topology  $\tau_P$ ).<sup>18</sup> In such cases, we may also say that model  $s$  has property  $p$  up to an  $\epsilon$ -error.

The paradigm of  $\epsilon$ -fidelity uses what will be called  $\delta$  to  $\epsilon$  deductions. Such deductions occur whenever it can be shown that a model that is  $\delta$ -close to having a particular property  $p$  must be  $\epsilon$ -close to having some further property  $q$ . More precisely, let  $\varphi : S \rightarrow P$  and  $\psi : S \rightarrow Q$  be property mappings from the space of mathematical models  $S$  to the respective property spaces  $P$  with a topology  $\tau_P$  and  $Q$  with a topology  $\tau_Q$  (see figure 2.2.1). Given the neighborhoods  $\delta(p) \subset P$  and  $\epsilon(q) \subset Q$  we may deduce that a model  $s \in S$  is  $\delta$ -close to having property  $p$  only if it is  $\epsilon$ -close to having property  $q$  whenever  $\varphi^{-1}[\delta(p) \cap \varphi[S]] \subset \psi^{-1}[\epsilon(q) \cap \psi[S]]$  as in figure 2.2.1. If in addition to this *mathematical* entailment, it is possible to determine by epistemologically legitimate abstraction methods that a given physical target is well matched to *some* model that is  $\delta$ -close to having property

---

<sup>17</sup>There may, of course, exist more than one topological structure defined on a given property space  $P$ . In such cases, there exist multiple modes of rigorously speaking of closeness relations with respect to properties in  $P$  so the relevant topology in use must be specified.

<sup>18</sup>Though the particular property mapping and topology on the property space is necessary for a well defined claim of  $\epsilon$ -closeness, when these are evident from the context without ambiguity, we will continue to leave these specifications implicit.



## 2.2. SOLVING PLATO'S PROBLEM

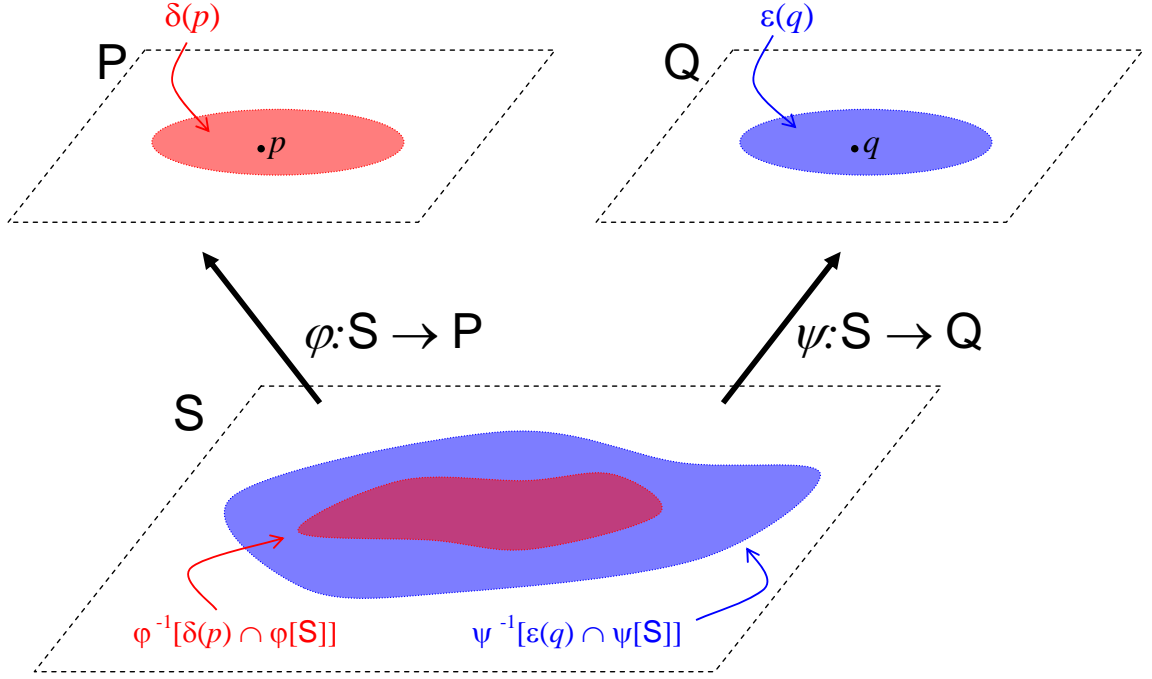


Figure 2.2.1: *Diagram of  $\delta$  to  $\epsilon$  deduction: any model in the set  $\varphi^{-1}[\delta(p) \cap \varphi[S]]$  of mathematical models in  $S$  homomorphically embedded by  $\varphi$  into the neighborhood  $\delta(p)$  of point  $p \in P$  is also in the set  $\psi^{-1}[\epsilon(q) \cap \psi[S]]$  of models embedded by  $\psi$  into the neighborhood  $\epsilon(q)$  of  $q \in Q$ .*

$p$ , then it will be epistemologically sound to deduce that the physical target is  $\epsilon$ -close to having property  $q$ . Hence, the representation is  $\epsilon$ -faithful.

This is precisely the form of deduction that was used on our carpentry example in section 2.1.3. In that example, both the  $P$  space and the  $Q$  space may be given by the ordered field of real numbers  $\mathbb{R}$ .<sup>19</sup> For the property space  $P$  we want the property assignment of an element  $\varphi(s) \in P$  to a quadrilateral  $s \in S$  to tell us the maximum difference between the side-lengths of  $s$  and the fixed value  $S$ . Hence we may define the  $P$ -property homomorphism  $\varphi$  as follows:

$$\varphi : s \mapsto \arg \max_{S_i} |S_i - S|$$

<sup>19</sup>Observe, the respective property spaces may be constituted by isomorphic, homomorphic, or (as in this case) identical mathematical spaces. What distinguishes the properties that they keep track of is their respective property homomorphisms.

## 2.2. SOLVING PLATO'S PROBLEM

where the  $S_i$  values range through the side lengths of the quadrilateral  $s \in \mathbf{S}$ . On the other hand, for the  $\mathbf{Q}$  space we want the property assignment of an element  $\psi(s) \in \mathbf{Q}$  to a quadrilateral  $s \in \mathbf{S}$  to tell us the maximum difference between the angle measurements of  $s$  and the fixed value  $\frac{\pi}{2}$ . Hence we may define the  $\mathbf{Q}$ -property homomorphism  $\psi$  as follows:

$$\psi : s \mapsto \arg \max_{\theta_i} \left| \theta_i - \frac{\pi}{2} \right|$$

where the  $\theta_i$  values range through the interior angle measures of the quadrilateral  $s \in \mathbf{S}$ .

Using the additive and ordering structures of the property spaces  $\mathbf{P}$  and  $\mathbf{Q}$  we also have a natural metric (and corresponding topological) structure defined on the respective spaces. Given these explicit specifications a model  $s \in \mathbf{S}$  can be defined as  $\delta$ -close to having the property  $S \in \mathbf{P}$ , whenever  $\varphi(s) \in \delta(S) := (S - \delta, S + \delta)$ . Hence, a set  $\mathcal{V}_{\delta,S}$  from section 2.1.3 is identical to the set of models in  $\mathbf{S}$  that get mapped to some point in the neighborhood  $(S - \delta, S + \delta)$ , i.e.  $\mathcal{V}_{\delta,S} = \varphi^{-1}[(S - \delta, S + \delta) \cap \varphi[\mathbf{S}]]$ . Similarly, a model  $s \in \mathbf{S}$  can now be defined as  $\epsilon$ -close to having the property  $\frac{\pi}{2} \in \mathbf{Q}$ , whenever  $\psi(s) \in \epsilon(\frac{\pi}{2}) := (\frac{\pi}{2} - \epsilon, \frac{\pi}{2} + \epsilon)$ .

By deducing the inequality  $\epsilon \leq \frac{\pi\delta(\delta+2)}{2(1-\delta)^2}$  in section 2.1.3 from law of cosines constraint imposed on all of our (Euclidean) models in  $\mathbf{S}$ ,<sup>20</sup> we were therefore able to take advantage of the respective metric structures preserved in the respective property spaces to conclude the containment relationship  $\varphi^{-1}[(S - \delta, S + \delta) \cap \varphi[\mathbf{S}]] \subset \psi^{-1}[(\frac{\pi}{2} - \epsilon, \frac{\pi}{2} + \epsilon) \cap \psi[\mathbf{S}]]$  on the models in  $\mathbf{S}$ . Hence, this representation counts as  $\epsilon$ -faithful, providing knowledge that a wooden physical target must be  $\epsilon$ -close to having right angles, whenever the carpenter can

---

<sup>20</sup>The reader may note that obeying the Euclidean geometry constraint of satisfying the law of cosines *perfectly* is as guilty of expecting perfect matching as the constraint of perfect congruence. As with our relaxation of the (perfectly) congruent diagonals constraint imposed on  $\mathbf{S}$  (see note 13 above), this constraint can be relaxed to allow for “ $\delta$ -imprecision” as well. That is to say, we may allow that the models (merely) be “almost Euclidean,” satisfying the law of cosines only up to some order of  $\delta$ . As in note 13, such a relaxation, though expanding membership in our new space of abstracted models  $\mathbf{S}'$ , ultimately allows for the deduction of a similar inequality.

## 2.2. SOLVING PLATO'S PROBLEM

gain the knowledge that the wooden frame is well matched to *some*  $\delta$ -close model  $s \in \mathbf{S}$ . Unlike the precise case, however, gaining knowledge that the carpenter's frame is " $\delta$ -close" in this way is far more epistemologically tractable than her gaining the precise knowledge required of matching accounts that are more narrow in scope.

It follows immediately that (perfectly) faithful representations also count as  $\epsilon$ -faithful representations. If we can *precisely* gain knowledge about a physical target by using a perfectly matched epistemically faithful representation, then such further knowledge is true up to *any*  $\epsilon$ -error, including no error. In this sense, an epistemically faithful representation is "trivially" an  $\epsilon$ -faithful representations since the  $\epsilon$ -error can be reduced to nothing. Though faithful representations are necessarily  $\epsilon$ -faithful, *not all  $\epsilon$ -faithful representations must be (strictly) faithful*. While (like the narrow scope representation of section 2.1.1) our "larger scope" representation of section 2.1.3 is not an epistemically faithful representation, it does count as an epistemically  $\epsilon$ -faithful representation, because it can be used to gain further knowledge of the physical target within a sufficiently small  $\epsilon$  margin of error.

Deductions from  $\delta$ -closeness to  $\epsilon$ -closeness to establish  $\epsilon$ -fidelity are especially robust when there exists the kind of  $\epsilon$ - $\delta$  deductive continuity demonstrated in the carpentry solution. To be precise,  $\epsilon$ - $\delta$  *deductive continuity* is said to exist with respect to a particular  $\mathbf{Q}$ -property,  $q$ , if for every sufficiently small  $\epsilon$ -error with respect to  $q$ , there exists some  $\delta$ -neighborhood in the topology of  $\mathbf{P}$  such that all such  $\delta$ -close models must be at least  $\epsilon$ -close to  $q$ . The existence of a  $\delta$  to  $\epsilon$  deduction establishing the  $\epsilon$ -fidelity of a representation does not entail the existence of deductive continuity. It may be that a representation is  $\epsilon$ -faithful only for deductions concerning certain fixed  $\epsilon$ -errors. However, the existence of such continuity can be far more epistemically robust, enabling not only knowledge about target systems for a *particular* sufficiently small error margin, but also establishing a range of potential margins

## 2.2. SOLVING PLATO'S PROBLEM

and the corresponding  $\delta$ -neighborhoods that the physical system must meet in order to remain within the respective  $\epsilon$ -error margins. Moreover, as we shall observe in section 2.3, such deductive continuity enables knowledge of a kind of “limiting behavior” about how narrowing the  $\delta$ -neighborhoods forces the Q-properties of such  $\delta$ -close models to approach the property  $q \in \mathbf{Q}$ .

An important point that will become relevant in section 2.3 is that  $\delta$  to  $\epsilon$  deductions establishing  $\epsilon$ -fidelity as we have defined them do not require the existence of a model  $s \in \mathbf{S}$  such that  $\varphi(s) = p$  and  $\psi(s) = q$ . Though the selection of specific points  $p$  and  $q$  play an essential role in “anchoring” the respective senses of  $\delta$  and  $\epsilon$  closeness in their respective  $\mathbf{P}$  and  $\mathbf{Q}$  property spaces, it is *not necessary* that any mathematical model of a representation actually satisfy these *anchor properties* (even in cases of deductive continuity). In the following, given a  $\delta$  to  $\epsilon$  deduction we will refer to the set  $\varphi^{-1}[p] \cap \psi^{-1}[q] \subset \mathbf{S}$  as the set of *anchor models* of the deduction.<sup>21</sup>

This language allows us to render precise the temptation (raised at the close of section 2.1.1) to refer to the mismatching of the rhombus with perfectly congruent sides and our physical target as an innocuous “idealization.” The idealization works because the idealized model was an anchor model of the  $\delta$  to  $\epsilon$  deduction developed in section 2.1.3. In fact, this anchor model was used to establish anchor properties for which we could establish  $\epsilon$ - $\delta$  deductive continuity. Hence, the role of this idealization as an anchor model is not only legitimate, but quite epistemologically robust in establishing the  $\epsilon$ -fidelity of our (wide scope) representation.

It is our proposal that many (if not most) successful (i.e. knowledge generating) applications of “idealized” properties or models can be shown to be successful because of the

---

<sup>21</sup>Again, since property mappings need not be surjective, it is possible that  $\psi^{-1}[q] = \emptyset$ , or  $\varphi^{-1}[p] = \emptyset$ .

## 2.2. SOLVING PLATO’S PROBLEM

existence of some  $\epsilon$ -faithful representation, taking advantage of (wide scope) extrinsic closeness relationships. The plausibility of this proposal is supported by the fact that *most scientific knowledge predicted by making use of a mathematical representation is only ever confirmed up to some suitably small degree of accuracy*. For example, if our carpenter wanted to confirm directly whether or not the angles of her wooden frame were actually right, she could always do so by direct measurement with a well calibrated woodworking protractor. However, even the best protractor can only be used to reliably measure the angle of a joint up to a certain degree of precision. Hence, even the most precise tools only ever confirm the knowledge gained by using a congruent diagonals type deduction up to a certain ( $\epsilon$ ) margin of error.

This is not unique to carpentry. All scientific observations are ultimately bounded by some level of precision beyond which our instruments cannot measure. If a given mathematical representation is used to gain some further knowledge about targeted physical phenomena, this new knowledge can only ever be confirmed up to this precision level. That is to say, for a sufficiently small margin of error  $\epsilon$  below the precision level of our measuring devices, *knowledge gained through an  $\epsilon$ -faithful representation is no less confirmed by direct observation than a “perfectly” faithful representation*. Unlike the latter case however, the  $\epsilon$ -fidelity means this representation enables knowledge without running into the soundness dilemma generated by Plato’s problem.

### 2.2.2 $\epsilon$ -Fidelity and Experimentation

The practical constraints of our instrumentation are not limited to the imprecision of confirming knowledge gained by mathematical representation (the “ $\epsilon$  end”). They also influence the degree of precision available in determining the initially matched properties attributed

## 2.2. SOLVING PLATO’S PROBLEM

to a physical phenomenon (the “ $\delta$  end”). Again, in the carpentry example Plato’s problem did not result so much from the fact that we *knew* that the interior edges of the physical wooden frame were not congruent. Rather, we could not be reasonably confident that they were *at every level of precision*.<sup>22</sup> Even if our carpenter were to measure each side with her most reliable measuring tape multiple times, she could not not confirm that the lengths are perfectly congruent at any scale of precision. What she can determine with reasonable confidence (from suitable measurement techniques and devices) is that the four edges of her wooden frame are congruent (of measure  $S$ ) *up to an error of  $\delta$* . That is to say, she *can* reasonably detect through measurement that the wooden frame has the kind of abstract, stable, phenomenal properties well matched with *some* member of  $\mathcal{V}_{\delta,S}$ .

Of course, the practical constraints of measurement precision prevent her from determining the *particular member* of  $\mathcal{V}_{\delta,S}$  to which the stable properties abstracted from her wooden frame are perfectly matched. However, one of the key benefits of the kind of  $\delta$  to  $\epsilon$  deduction used to establish  $\epsilon$ -fidelity is that knowing the particular member is not required. All we need in order to make use of the deductions for such an  $\epsilon$ -faithful representation is the ability to detect that it is well matched with *some* member, even if we cannot determine which member in particular.

This is one example of how understanding the  $\epsilon$ -fidelity of a representation helps to eliminate some of the “epistemic debt” incurred in abstraction processes. One of the sources discussed in section 1.3 was that in extracting and aggregating the data from the physical

---

<sup>22</sup>Note, the claim being made is *in no way* a claim of scientific anti-realism or skepticism. The claim here merely pertains to the practical constraint that our measuring devices do not provide unlimited precision. While a motivation for developing an account of  $\epsilon$ -faithful representation is that we cannot assume unlimited precision to justify knowledge gained (in part) through mathematical representation, a philosopher of science can consistently make a claim about the (abductive?) theoretical justification for the existence of an unobservable entity (e.g. a photon) while still recognizing that certain knowledge about the phenomenal properties of the entity are justified *only up to some  $\epsilon$ -error*.

## 2.2. SOLVING PLATO'S PROBLEM

system our measurement procedures are subject to experimental noise. Imagine we wish to determine the value  $X$  of a measurable property of a physical system, or system type. We have a procedure for coming up with measurement values  $X_i$  but due to the noise, these values are not always identical. If the effects of noise sources meet certain conditions such as being independent, accumulating linearly, and having symmetrical influence, then the average value of such noise sources should converge to nothing with repeated measurements. That is to say, if we measure the value of  $X$  enough times, then the average after  $n$  trials  $\bar{X}(n) := \frac{1}{n} \sum X_i$  will likely approach the obfuscated value  $X$ . Averaging such repeated measurements is one simple example of how we might use data to detect a stable phenomenal property. Though it may not be true that  $\bar{X}(n) = X$  for any particular finite number of measurements  $n$ , for large enough  $n$  the variation of the potential  $\bar{X}(n)$  results decreases with order  $\frac{1}{n}$ , which in turn allows us to increase our confidence that our actual  $\bar{X}(n)$  result is  $\delta$ -close to the value  $X$ . Our confidence that the aggregated value  $\bar{X}(n)$  is close to  $X$  can be made as large as we like by taking a sufficiently large number of measurements.

To be clear, because we cannot *completely* rule out the possibility of certain (very unlikely) outlier possibilities we cannot say that the value  $|\bar{X}(n) - X|$  *necessarily* goes to 0. However, we can say our *confidence* that they are close can be made arbitrarily large (i.e. for any  $\hat{\delta} > 0$ , the probability  $P(|\bar{X}(n) - X| > \hat{\delta}) \rightarrow 0$  as  $\frac{1}{n} \rightarrow 0$ ).<sup>23</sup> Note also that the relationship between the  $\frac{1}{n}$  values and the  $\hat{\delta}$  values here is (somewhat) analogous to the type of  $\epsilon$ - $\delta$  continuity relationship discussed above, where the  $\frac{1}{n}$  values play the condition or “ $\delta$  role,” and the probability measurable sets associated with the  $\hat{\delta}$  values play the error or “ $\epsilon$  role.” The idea is that, if with enough measurements (i.e. small enough  $\frac{1}{n}$ ) we sufficiently increase our confidence that  $|\bar{X}(n) - X| < \hat{\delta}$ , then we can now *justifiably* use our confidence that

---

<sup>23</sup>This is the weak version of the law of large numbers.

## 2.2. SOLVING PLATO'S PROBLEM

$\bar{X}(n)$  is within a  $\hat{\delta}$ -neighborhood of  $X$  to potentially draw new conclusions about the system up to an  $\epsilon$ -error.

In contrast, requiring a perfect matching between a model with a *precise* value of  $\bar{X}(n)$  and the physical system (which *ex hypothesi* has a value of  $X$ ) in order to make a deduction is epistemologically highly suspect. That is to say, we can only have a vanishing small confidence that  $|\bar{X}(n) - X| = 0$  *exactly* even with an arbitrarily large number  $n$  of repeated measurements. In the case of  $\epsilon$ -fidelity, however, when our conclusion only requires some level of “ $\delta$ -closeness” in matching between the model with a value of  $\bar{X}(n)$  and a physical system (in order to deduce some further fact about the system within an  $\epsilon$  margin of error), then given enough measurements *we can be very confident* that the physical system is well matched to our model or one of its  $\delta$ -close neighbors. Hence, by accommodating for the “ $\delta$ -imprecision” in our measurement and then detection of the stable phenomenal properties, we are able to attribute to a system (or system type) an  $\epsilon$ -faithful representation that can eliminate some of the epistemic debt incurred at the measurement and detection stage of abstraction.

### 2.2.3 Putting the $\epsilon$ in Subtle Transition

Our account of  $\epsilon$ -fidelity can also be used to resolve some of the representational difficulties that occur when different theoretical presuppositions (epistemic debts) come into conflict. Recall, according to thermodynamics, the stable (phenomenal) relationship between state variables (like pressure and volume) exhibits non-analyticities during phase transitions. This results in kinks or discontinuities in functions in these variables (see e.g. figure 1.3.1). In the theoretical context of thermodynamics the data is fit piecewise to allow for this non-analytic behavior at the phase transition boundaries. In contrast, in statistical mechanics of systems



## 2.2. SOLVING PLATO’S PROBLEM

with a finite number of degrees of freedom (e.g. finite number of particles) it is not possible to recover such non-analytic relationships between the thermodynamic state variables like pressure, temperature and volume. In statistical mechanics, as in thermodynamics, variables like these are identified with partial derivatives of a free energy function. However, unlike thermodynamics, in statistical mechanics free energy ( $F$ ) is calculated with the following following equation:

$$F = -kT \cdot \log(Z),$$

where  $k$  is Boltzmann’s constant,  $T$  is temperature and,  $Z$  is the partition function calculated by taking the following sum:

$$Z = \sum_c e^{\left(\frac{-H(c)}{kT}\right)}$$

where  $H(c)$  is the energy of the system in configuration  $c$ .

The difficulty is that in statistical mechanics the partition function of a system with a *finite number* ( $N < \infty$ ) of degrees of freedom can only have a finite number of configurations  $c$ . So the partition function is calculated by taking finite sums of positive exponentials. Hence, for finite systems, the partition function, the free energy function, and any partial derivatives of these functions must be analytic.<sup>24</sup> The mathematical fact that finite system models of statistical mechanics cannot exhibit non-analyticities has led physicists such as Leo Kadanoff to make seemingly strange sounding claims like “[s]ince phase transitions only happen in an infinite system, we cannot say that any phase transitions actually occur in the finite objects that appear in our world” (Kadanoff, 2009, p10).<sup>25</sup> He goes on to explain

---

<sup>24</sup>See e.g. (Liu, 1999) for nice technical elaboration of this piece of the argument. See also (Callender, 2001, Batterman, 2002).

<sup>25</sup>See also (Kadanoff, 2000, p238).

## 2.2. SOLVING PLATO’S PROBLEM

that “[n]ature gives us no pure thermodynamic phases but only real objects displaying their own complex and messy behavior.” Kadanoff’s ostensible argument for the claim is as follows: According to statistical mechanics, the non-analyticities precisely defining “pure thermodynamic phase” boundaries cannot be modeled in finite systems. Precisely defined phase boundaries are necessary for the existence of phase transitions. Therefore, according to statistical mechanics, no “pure thermodynamic phase” boundaries exist through which systems can transition.

Since statistical mechanics is supposed to be our “more fundamental” thermal theory, Kadanoff presumably draws his unqualified claim (about the non-existence of (pure) thermodynamic phases “in our world”) from this conclusion. The specification that he is actually only denying the existence of *pure thermodynamic phases* takes some of the sting out of Kadanoff’s claim: Finite ice cubes can still melt according to statistical mechanics and solid ice is still quite distinguishable from liquid water (in most thermal states). What Kadanoff is actually arguing for is that the *precise phase boundaries*, occurring in the regions of thermodynamic state space where non-analytic kinks and discontinuities are identified, cannot be pinpointed with finite models of statistical mechanics. The transitions are more subtle.

Part of the difficulty with this example has to do with the fact that conflicting theoretical presuppositions can influence how we abstract the phenomenal relationships from the data.<sup>26</sup> As already emphasized, in thermodynamics this can be done piecewise where the pieces need not be joined analytically. The relevant state relationships can hence be matched to thermodynamic models with “pure” phases meeting at discrete boundaries. In the context of finite system statistical mechanics, this sort of relationship is not available. What is available are subtle transitions characterized by very “steep” (but continuous) changes in state variables

---

<sup>26</sup>Cf. Bangu’s discussion on this topic in (Bangu, 2009).

## 2.2. SOLVING PLATO’S PROBLEM

or very rapid (but unknicked) changes in their slope, etc. These kinds of behavior are *similar* in some respects to the discrete behavior modeled in the thermodynamic cases, but they are not identical.

If we want to *exactly* match a statistical mechanics model to the discretely bounded phenomenal behavior that we might abstract in the thermodynamic context, then only infinite  $N$  statistical mechanics models will work. If we give up such an exact matching requirement, then given the right kind of  $\epsilon$ -closeness (see section 2.3.2 below), it may be possible to find statistical mechanics models that fall within the appropriate neighborhoods of models that do have the (mathematical properties) of models exhibiting what Kadanoff calls *pure* thermodynamic phase boundaries but which need only a finite number of degrees of freedom themselves. If such  $\epsilon$ -neighborhoods can be identified, then despite Kadanoff’s point we may use  $\epsilon$ -faithful representations to gain knowledge of systems with “their own complex and messy behavior” whenever they are suitably ( $\delta$ ) close in the right way to crisply behaved but idealized anchor models. This will allow us to use mathematical representations to understand and gain knowledge of even the messy systems with their subtle (and never actually discrete) phase transitions.

In the next section we will take a closer look at how this can be done even when the idealized model is forced to meet certain unrealistic constraints like having an infinite particle number, but, before closing this section, it is worth observing how such an  $\epsilon$ -faithful representation of phase transitions discharges another kind of the epistemic debt incurred through the abstraction process. As we have mentioned, abstracting from data like that of figure 1.3.2 depends on theoretical presupposition: if it is thermodynamics, we can choose from the space of functions with non-analyticities, whereas with (finite) statistical mechanics perhaps only analytic functions should be allowed. Once we shift our attention to the wider

## 2.2. SOLVING PLATO’S PROBLEM

context matching of  $\epsilon$ -faithful representation, these restrictions can be relaxed (i.e. the space of abstracted models  $S$  may include models of both types). We might use an “idealized” anchor model with determinate kinks or discontinuities indicating in our model the boundaries of “pure thermodynamic phases.” However, in the case of  $\epsilon$ -faithful representation such an idealized anchor model need not be directly matched to the physical target in every (or any) respect.

It is the physical target and its extrinsic relationships to other potential physical systems that is matched by the closeness relations characterizing the appropriate topological proximity to this ideal system. Again, while in this sense the ideal system *anchors* the relevant neighborhoods that are ( $\epsilon$ - or  $\delta$ -)close to it, we need not take an epistemic position about the particular neighboring member to which it is best matched. Hence, we need not take on the epistemic debt of insisting that phenomenal relationships can only be extracted from the data in the form of analytic or non-analytic functions exclusively. The  $\delta$ -neighborhoods may include both analytic and non-analytic models. So as long as they are  $\delta$ -close to the anchor model in a way that lets us know that they must be  $\epsilon$ -close to the anchor property (of the idealized anchor model, if it exists), the debt of such further theoretical presupposition can be avoided.<sup>27</sup>

---

<sup>27</sup>The example of note 20 in which we considered admitting non-Euclidean quadrilaterals into our space and then restricted our  $\delta$  neighborhoods to include only models with a sufficiently small (order  $\delta$ ) deviation from the law of cosines is another example of this. Such a move discharges the theoretical presupposition that physical objects (even on earth) must obey the geometric relations of Euclidean geometry *perfectly*. Such a representation may hence avoid the epistemic debt possibly incurred by committing exclusively to Newtonian presuppositions over more relativistic ones.

## 2.3 $\epsilon$ -Fidelity and Proximity to Pathology

In section 2.2.1 we developed our account of  $\epsilon$ -faithful representation, explaining how such representations simultaneously resolve both the mysterious fidelity problem and Plato’s problem by making use of  $\delta$  to  $\epsilon$  deductions that can accommodate the experimental imprecision and ampliative presuppositions inevitable in the abstraction of scientific phenomena. In outlining how an  $\epsilon$ -faithful representation of finite thermal systems might be used to understand or gain knowledge about the first-order phase transitions of such systems we proposed possibly letting models with an infinite number of degrees of freedom serve as the anchor models for our  $\delta$  to  $\epsilon$  deductions. In this section we will now consider what happens when such “pathological idealizations” play the function of anchor models, where a model counts as *pathological* whenever it meets constraints that either (I) cannot possibly be matched by any physical system or (II) render the mathematical model incompatible with the defining constraints of the relevant physical theory. This investigation will complete our account of  $\epsilon$ -faithful representation, demonstrating that, as with non-pathological idealizations, appropriately using only proximity to a pathological anchor models avoids the epistemological complications of matching the idealized model while continuing to enable the advantages argued for above.

### 2.3.1 Limits and Anchor Properties

To motivate the arguments of the remainder of this chapter, in this section we will consider a toy example based on the famous “halving a square” challenge from Plato’s *Meno*. Imagine that there exists a fictitious substance called *rationalium* with the following properties: Rationallium consists of discrete cube shaped “atoms” with side lengths of exactly 1 *unit*.

### 2.3. $\epsilon$ -FIDELITY AND PROXIMITY TO PATHOLOGY

Imagine now that Plato asks us to consider pairs of “square shaped” stacks of rationallium atoms exactly 1 rationallium atom deep, and stacked  $n$  atoms wide and tall for some  $n \in \mathbb{N}^+$ . The challenge is to find a particular pair of these “square shaped” rationallium stacks such that the volume of the larger stack is exactly twice the volume of the smaller stack.

Observe that though our challenge is similar to the one posed to the servant from the *Meno* (in that for both challenges we are looking for “square shaped” constructions with a ratio of 2 to 1), the original challenge was an exercise in mathematical deduction. In contrast, our challenge is one of mathematical application to the “physical” (albeit fictitious) rationallium systems. The subtlety of this difference between the two challenges is made clear by the fact that unlike the *Meno* case (which, so the story goes, is solvable even by agents with no mathematical training), our challenge has no solution. It is impossible to construct two such square shaped stacks out of pure rationallium with a volume ratio of exactly 2 to 1.

The essential reason why the mathematical *application* challenge fails where the pure mathematics challenge succeeds follows from the irrationality of the number  $\sqrt{2}$ . Recall, the solution in the case of the *Meno* is to “quarter” the larger square and then construct the smaller square from the diagonals of the quarters (see figure 2.3.1). So if the side length of the larger square is 8 then the diagonal lengths of its quarters (and so the side lengths of the smaller square) must be  $4\sqrt{2}$ . Though allegedly mystifying to the ancient Pythagoreans, such an abstract *mathematical construction* of a figure with side length  $4\sqrt{2}$  is entirely possible.

In contrast, the *physical construction* of “square shaped” stack of rationallium is not possible because rationallium only comes in fixed discrete units. For example, in figure 2.3.2 we have a  $1 \times 8 \times 8$  configuration for a total of exactly 64 atoms. So in order to construct

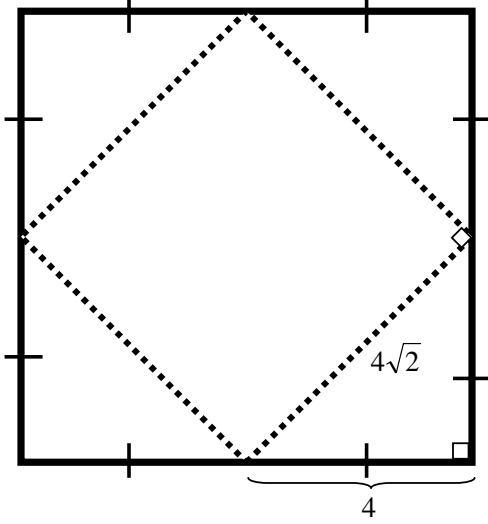


Figure 2.3.1: *Diagram of the solution to Plato's original challenge.*

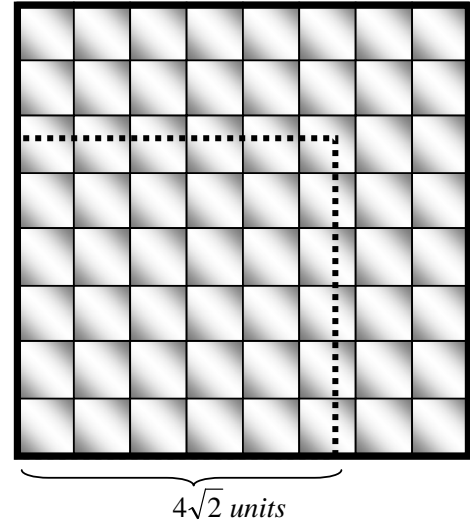


Figure 2.3.2: *Failed attempt to fit enough rationallium within a  $4\sqrt{2} \times 4\sqrt{2}$  square region suggested by the original solution.*

a configuration of exactly half the size, we need our smaller square configuration to consist of exactly 32 atoms. The problem, however, is that the square root of 32 is the *irrational* number  $4\sqrt{2}$ . The arithmetic generalizes for any  $1 \times n \times n$  model of our square shaped rationallium stacks. If the smaller square has dimensions  $1 \times m \times m$ , then the integers  $n$  and  $m$  would have to satisfy the relation  $n^2 = 2m^2$  or  $\frac{n}{m} = \sqrt{2}$ . But this is an impossible task. There are no rationallium atoms of irrational width. Since the numbers  $n$  and  $m$  must be integers  $\frac{n}{m}$  cannot be an irrational number. So, in contrast to the mathematical construction, it is impossible to construct a configuration of complete rationallium atoms with a ratio of exactly 2 to 1.

Though the above deduction is valid, there is a trick that we might attempt in order to meet the rationallium challenge. Noting that the decimal expansion of the  $\sqrt{2} = 1.41421\dots$ , we might construct a *sequence*, denoted by  $((a_i, b_i))_{i \in \mathbb{N}}$ , of pairs of “square shaped” rationallium stacks such that the larger square has the dimensions  $1 \times a_i \times a_i$  and the smaller square

### 2.3. $\epsilon$ -FIDELITY AND PROXIMITY TO PATHOLOGY

has dimensions  $1 \times b_i \times b_i$  where  $a_i := 2 \cdot 10^i$  and  $b_i := \lfloor 10^i \cdot \sqrt{2} \rfloor$  for each  $i \in \mathbb{N}$  (i.e. the sequence  $((a_i, b_i))_{i \in \mathbb{N}}$  is given by  $(2, 1), (20, 14), (200, 141), \dots$ ). If we look at this sequence far enough out, we will see that the sequence of ratios  $\left(\frac{a_i}{b_i}\right)_{i \in \mathbb{N}}$  of the larger side lengths to the smaller ones comes arbitrarily close to the number  $\sqrt{2}$ . In other words, in the limit as  $i \rightarrow \infty$  we have that the ratios of the lengths  $\frac{a_i}{b_i} \rightarrow \sqrt{2}$ . Hence, “in the limit” the ratio of the volumes  $\frac{1 \times a_i \times a_i}{1 \times b_i \times b_i} = \left(\frac{a_i}{b_i}\right)^2$  approaches 2.

Such analysis might be taken to suggest the rather bizarre claim that “if we were to make both of the squares ‘infinitely large’ in a particular way we could get a pair of rationallium stacks with a volume ratio of *exactly* 2 to 1!” Even if we did think that such a “limit pair” met Plato’s challenge in a sense, there is something very pathological about this solution: Beyond practical issues of getting our hands on the requisite infinite amounts of time, space, and rationallium, though the above *mathematical* analysis of the limiting behavior is sound, the “limit system” consisting of two rationallium configurations of infinite side length is not well defined.

Though we must conclude that no finite pairs of rationallium configurations can have a ratio of *exactly* 2 to 1, the convergence to  $\sqrt{2}$  of the ratios  $\frac{a_i}{b_i}$  does enable further knowledge about finite pairs of rationallium stacks. We cannot meet the rationallium challenge exactly, but if we broaden our focus to look for configurations with a ratio of *nearly* 2 to 1, we can find pairs that come *arbitrarily close* to having the right ratio. If we make our pairs large enough, our error in meeting Plato’s rationallium challenge, though never vanishing, can be made negligibly small. In this sense, the volume ratio  $\left(\frac{a_i}{b_i}\right)^2$  of the corresponding sufficiently large pairs of rationallium stacks can be said to *cluster in the neighborhood* of 2 to 1 even though none of them will ever meet the condition exactly.<sup>28</sup>

---

<sup>28</sup>This is an example of the kind of case described by Norton (2012) in which there exists an approximation



### 2.3. $\epsilon$ -FIDELITY AND PROXIMITY TO PATHOLOGY

Our account of  $\epsilon$ -faithful representation developed in section 2.2.1 allows us to understand how taking limits enable knowledge even in cases that the model is pathological or even non-existent in  $\mathbf{S}$ . Our space  $\mathbf{S}$  of models, denoted by ordered pairs of positive natural numbers  $(n, m)$ , consists of pairs of “square shaped” rationallium stacks with dimensions  $1 \times n \times n$  and  $1 \times m \times m$  respectively. The property space  $\mathbf{P}$  will be given by the interval  $[-1, 1]$  equipped with the overlapping interval topology. Our  $\mathbf{P}$ -property homomorphism  $\varphi$  can then be defined by sending the elements  $(a_i, b_i)$  defined above to the point  $\frac{1}{i+1} \in [-1, 1]$ , and sending all other  $(n, m)$  pairs to the point  $-1$ .<sup>29</sup> The overlapping interval topology on this property space in particular allows us identify the neighborhoods  $\delta(0) := (-\delta, \delta)$  around the point  $0 \in [-1, 1]$ . Next, we want our property space  $\mathbf{Q}$  to indicate the volume ratio of elements in  $\mathbf{S}$ . So identifying this property space with the real numbers  $\mathbb{R}$  equipped with the natural topology, an appropriate  $\mathbf{Q}$ -property homomorphism  $\psi$  can then be defined by the operation  $\psi : (n, m) \mapsto (\frac{n}{m})^2$ . The limit discussed above can now be stated in terms the resulting (continuous)  $\epsilon$ - $\delta$  deductions: for every  $\epsilon$ -error (viz every neighborhood  $(2 - \epsilon, 2 + \epsilon)$  around the exact value of 2 in the volume ratios property space  $\mathbf{Q}$ ), there exists some  $\delta$ -neighborhood such that all models in  $\mathbf{S}$  that are  $\delta$ -close to 0 must be  $\epsilon$ -close to having a volume ratio of 2. Hence, for any value  $\epsilon$ , if we want a pair of rationallium atoms with a ratio of 2, up to that  $\epsilon$ -error, we know that there are pairs of rationallium squares  $(a_i, b_i)$  that come that close or closer.

Observe, despite the existence of this convergence relationship, and the knowledge we can gain from such an  $\epsilon$ -faithful representation, there does not exist any anchor model  $s \in \mathbf{S}$

---

but (because there is no limit system) it fails to count as an idealization.

<sup>29</sup>Note, because all we are really interested in for this example is tracking how far along an element  $(a_i, b_i)$  is in the sequence, if it is in the sequence, numerous other formulations would also suffice as a  $\mathbf{P}$  property space and its corresponding property homomorphism.

### 2.3. $\epsilon$ -FIDELITY AND PROXIMITY TO PATHOLOGY

such that  $\varphi(s) = 0$  and  $\psi(s) = 2$ .<sup>30</sup> Such an anchor model would constitute the kind of “limit system” of rationallium stacks of infinite side lengths ruled out above. Nevertheless, as emphasized at the close of section 2.2.1, despite the absence of such an anchor model, the existence of the  $\delta$  to  $\epsilon$  deductions alone enables the knowledge gained by our ( $\epsilon$ -faithful) representation.

This observation will become highly relevant to our discussion in the remainder of this chapter. The reason for this is that our analysis of the  $\delta$  to  $\epsilon$  deductions about phase transition phenomena will involve anchor models that qualify as *pathological* idealizations. Though we *may* be able to abstract a model space  $\mathbf{S}$  containing appropriate anchor models, exhibiting what Kadanoff refers to as “pure phase transitions,” because they are forced to meet constraints such as having an infinite number of degrees of freedom, arguments can be made for why they could not possibly be matched with any physical system.

The moral of this section is that such arguments do not matter for  $\epsilon$ -fidelity. Even if such anchor models only exist as (merely) inductive extrapolations of intrinsic and extrinsic relations possessed by actual (finite) physical systems, as we saw in this section, the existence of an anchor model in our space of abstract models  $\mathbf{S}$  is not required for us to gain knowledge using  $\epsilon$ -faithful representations. When understood primarily as exhibits of the *anchor properties* that are in fact essential for our  $\delta$  to  $\epsilon$  deductions, anchor models can be illustrative even when pathological. However, as our rationallium example demonstrates,  $\epsilon$ -fidelity can be established even when no anchor model exists. Hence, including a particular anchor model in  $\mathbf{S}$  at the (ampliative) abstraction stage is not strictly necessary for the epistemic gains made though  $\epsilon$ -fidelity.<sup>31</sup>

---

<sup>30</sup>The technical reason for this is that the sequences  $(a_i)_{i \in \mathbb{N}}$  and  $(b_i)_{i \in \mathbb{N}}$  both become arbitrarily large as  $i \rightarrow \infty$ , entailing that  $\lim_{i \rightarrow \infty} (a_i, b_i)$  does not converge to an element in  $\mathbb{N}^2$  and hence is not in  $\mathbf{S}$ .

<sup>31</sup>This is yet another example of how  $\epsilon$ -fidelity allows us to eliminate some of the epistemic debt possibly

#### 2.3.2 Closeness in Context

Non-trivially  $\epsilon$ -faithful representations do not enable perfectly precise knowledge of a physical target. Instead these representations enable us to infer that the system (or system type) must be *close* to having a particular property. So, it is important, in any case of non-trivial  $\epsilon$ -fidelity that we understand the physical relevance of the particular kind of closeness invoked.

In our carpentry example the relevant kind of  $\epsilon$ -closeness took advantage of the natural topology of the real numbers. Angle measurements attributed to physical wood joints are given real number values. Moreover, (to the degree of precision we can observe) the more similar the angles of two physical joints, the closer our angle measurements are numerically (i.e. in the topology of the real numbers). In other words, (at least at observable scales) the topological structure of the real numbers nicely keeps track of these physical similarities abstracted from wooden joints. For this reason the topological structure of the real numbers is appropriate for the  $\epsilon$ -closeness in our carpentry example. The chosen topology on the property space  $\mathbf{Q}$  keeps track of the kind of closeness in physical properties that we want to know about. A very similar story exists for identifying the  $\delta$ -closeness used in our carpentry example. Because the topological structure of the real numbers similarly keeps track of closeness in the length of wood beams, the neighborhood relations embedded in the real number topology was appropriate for the sense of  $\delta$ -closeness used in the example.

To illustrate the importance of employing a property space equipped with a topology that appropriately tracks the sense of closeness in physical properties we wish to know about, let us return to our discussion of  $\epsilon$ -faithful representations of first-order phase transitions from section 2.2.3. Recall, the only way to recover what Kadanoff referred to as “pure thermodynamic phases” from a model in statistical mechanics (e.g. the partition function of

---

incurred through abstraction.

### 2.3. $\epsilon$ -FIDELITY AND PROXIMITY TO PATHOLOGY

an abstraction of a thermal system such as an Ising model) was by having an infinite number of degrees of freedom in the system. Partition functions of systems with finite degrees of freedom are necessarily analytic, and hence cannot exhibit the non-analyticities used to pinpoint thermodynamic phase boundaries.

The dichotomy between models that can be non-analytic (i.e. models with  $N = \infty$  degrees of freedom) and those that cannot (i.e. models with  $N < \infty$ ) has led Robert Batterman (2010, p18) to suggest that such necessary conditions for non-analytic behavior “make it impossible to tell any kind of de-idealizing story that would enable one to rank idealizations in terms of their distance from a matching model (Pincock) or from full isomorphism (Bueno and Colyvan).” Batterman is clearly correct that under the kind of “narrow” matching accounts championed by Pincock and by Bueno and Colyvan, it does not appear possible to develop an epistemologically coherent account for how such models may be “de-idealized.”<sup>32</sup> If other closeness relationships can be included in a “wide scope” matching account though, the situation can change. When the right kind of extrinsic closeness relationships are also matched, it may be possible to use such closeness relationships to establish the  $\epsilon$ -fidelity of a representation, even in cases like phase transitions where the anchor model in question is so qualitatively distinct (in terms of its infinite degrees of freedom and its non-analytic behavior) from its  $\epsilon$ -neighbors.

Let us consider Batterman’s argument for why non-analytic statistical mechanics models (i.e. ones that must have  $N = \infty$  degrees of freedom) are not “de-idealizable.” One reason is that the thermodynamic limit constitutes what Batterman calls a “singular limit” defined by sequences where “the behavior as one approaches the limit is qualitatively different from the behavior one would have at the limit.” The qualitative distinction between the relevant

---

<sup>32</sup>See note 6 above.

### 2.3. $\epsilon$ -FIDELITY AND PROXIMITY TO PATHOLOGY

$N = \infty$  models in statistical mechanics and the  $N < \infty$  models is that the former can be non-analytic whereas the latter cannot:

Finite systems with more and more particles may in some sense get ‘close to’ the nonanalytic behavior in the thermodynamic functions, but for finite  $N$ , the curves are always smooth. There is no distance measure or metric saying how close an actually smooth curve is to a nonsmooth/nonanalytic one. (Batterman, 2010, p18)

Considered out of context, this second sentence may seem problematic. After all, there are in fact an infinite number of metrics that can be defined to measure how close a smooth curve is to behaving like a non-smooth or non-analytic curve in a given compactly contained region.<sup>33</sup> Moreover, from each of these metrics a corresponding topology may be defined to keep track of how close a particular smooth curve comes (in that topology) to a non-analytic curve, and these topologies may be put to good use in drawing  $\delta$  to  $\epsilon$  deductions about the kinds of (finite) models exhibiting such smooth behavior when in (topological) proximity to a non-analytic and hence infinite model.

Batterman elaborates his point about the dichotomy with these other modes of tracking closeness in behavior in a footnote to the above long quote:

For instance, the relevant curves ‘look’ sharper and sharper as the number of particles increases. But ‘looking sharp’ is not a relevant measure: For any finite  $N$ , no matter how large, the curves are smooth and analytic, no matter how sharp they appear. (Batterman, 2010, p18)

To understand what is going on let us consider an example of first-order phase transitions of ferromagnetic Ising models.<sup>34</sup> In figure 2.3.3 we have an illustration of the magnetization

---

<sup>33</sup>For example, the metric structure inherited from any  $\mathcal{L}^p(\Omega)$  norm taken over the compactly contained region  $\Omega$  in the domain of the smooth curves may be used to compare any smooth curve with a non-analytic curve that is e.g. suitably integrable or essentially bounded on  $\Omega$ . If the non-analytic curve meets the appropriate weak differentiability conditions, certain Sobolev norms may also be used to track closeness with respect to various differential properties of interest.

<sup>34</sup>For a similar discussion of this particular example see (Kadanoff, 2009, §1). Note, following Kadanoff the

### 2.3. $\epsilon$ -FIDELITY AND PROXIMITY TO PATHOLOGY

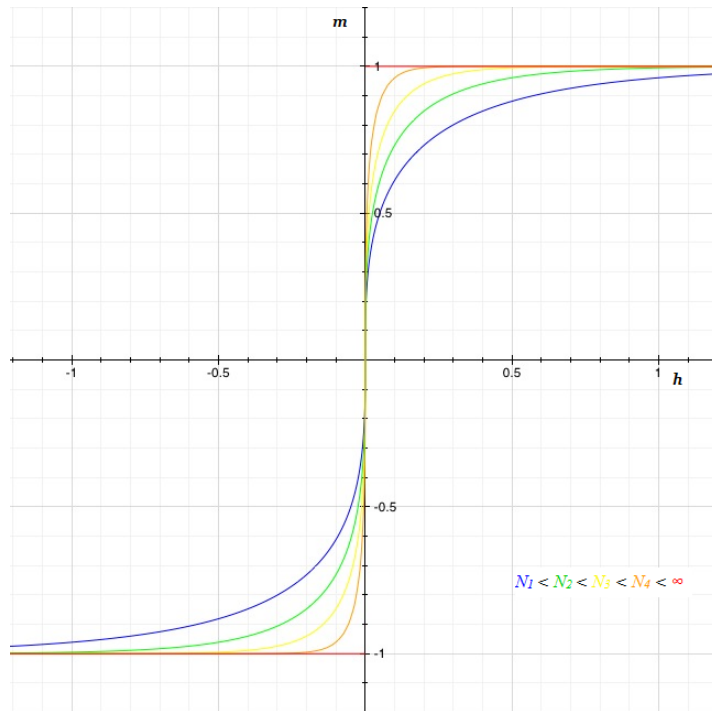


Figure 2.3.3: Graph of magnetization density curves  $m_N(h)$  as a function of the magnetic field  $h$  for systems with respectively increasing particle numbers. The absolute area enclosed by a curve  $m_N(h)$  and the discontinuous curve  $m_\infty(h)$ , depicted in red, becomes negligible as  $N \rightarrow \infty$ .

density  $m_N := \frac{M}{\mu N}$  as a function of the magnetic field  $h$  for successively increasing values of  $N$ . As the degrees of freedom  $N$  increase, the corresponding curves  $m_N(h)$  appear to be getting “closer and closer” to the discontinuous “limit curve”  $m_\infty(h)$  corresponding to the case when  $N = \infty$ . In fact, we can define a precise metric quantifying how “close” each of our finite curves  $m_N(h)$  comes to the limit curve  $m_\infty(h)$  over any bounded interval  $\Omega$  of  $h$  values with the  $\mathcal{L}^1(\Omega)$  norm:

---

graphics depicted in figure 2.3.3 should only be taken as “cartoon views.” Since our purpose here is (merely) to illustrate the available notions of continuous convergence in this context, we will not present the relevant finite-size scaling and simulation methods typically employed to generate such curve families. We will also be ignoring issues associated with metastability and observation times. Results pertaining to convergence in the simplest case of first-order ferromagnetic phenomena depicted here can be found in (Fisher & Berker, 1982, Binder & Landau, 1984). For a contemporary example of the application of these techniques to more sophisticated phase transition phenomena see e.g. (Zhou *et al.*, 2008). See also (Landau & Binder, 2005).

### 2.3. $\epsilon$ -FIDELITY AND PROXIMITY TO PATHOLOGY

$$d_{\mathcal{L}^1(\Omega)} : (m_N(h), m_\infty(h)) \mapsto \|m_N(h) - m_\infty(h)\|_{\mathcal{L}^1(\Omega)} \quad (2.3.1)$$

where,

$$\|f(x)\|_{\mathcal{L}^1(\Omega)} := \int_{\Omega} |f(x)| dx. \quad (2.3.2)$$

This metric can quantify the cumulative absolute difference between a curve  $m_N(h)$  and the limit curve over any such selected region  $\Omega$ . Hence, the topology  $\tau_{\mathcal{L}^1(\Omega)}$  inherited from the metric  $d_{\mathcal{L}^1(\Omega)}$  offers at least one physically relevant definition of how close the various magnetization curves come to behaving “like” the non-analytic curve as  $N \rightarrow \infty$ .

Moreover, there appears to exist an  $\epsilon$ - $\delta$  deductive continuity relationship with respect to  $\tau_{\mathcal{L}^1(\Omega)}$  neighborhoods of the infinite model. Let  $\mathbf{S}_{N \leq \infty}^d$  be the space of all  $d$ -dimensional Ising models with  $N \leq \infty$  degrees of freedom.<sup>35</sup> For our space of P-properties, we will again use the interval  $[-1, 1]$  equipped with the overlapping interval topology, and we will define the P-property homomorphism by  $\varphi : s_N \mapsto \frac{1}{1+N}$  where  $N$  is the degrees of freedom of a given model  $s_N \in \mathbf{S}_{N \leq \infty}^d$ . For our Q-properties, we may use the space  $\mathcal{L}^1(\Omega)$  of all integrable functions defined on an appropriate  $\Omega$  interval and equipped with the metric structure  $(d_{\mathcal{L}^1(\Omega)})$  and topological structure  $(\tau_{\mathcal{L}^1(\Omega)})$  induced by the  $\mathcal{L}^1(\Omega)$  norm according to the equations (2.3.1) and (2.3.2).<sup>36</sup> Finally, the Q-property mapping  $\psi : s_N \mapsto m_N(h)|_{\Omega}$  sends a given model  $s_N \in \mathbf{S}_{N \leq \infty}^d$  to its magnetization density function restricted over  $\Omega$ .<sup>37</sup>

---

<sup>35</sup>For the purposes of our current discussion, we will ignore here the (otherwise philosophically significant) question of whether this space of models is a legitimate abstraction of the relevant intrinsic and extrinsic relations detectible through observation of actual thermal systems.

<sup>36</sup>The interval  $\Omega$  will signify the range of field values  $h$  over which we wish to compare the various magnetization curves. Though this interval can be as large (or small) as we like and made to include the value  $h = 0$  where an ostensible transition occurs, we keep it bounded to ensure individual integrability.

<sup>37</sup>Recall, functional dependencies such as  $m_N(h)$  are found by taking the appropriate partial derivatives of the free energy, determined in statistical mechanics through the partition function of  $s_N$ .

### 2.3. $\epsilon$ -FIDELITY AND PROXIMITY TO PATHOLOGY

So *one way* of specifying the claim that as  $N \rightarrow \infty$  the magnetization density curves  $m_N(h)$  come arbitrarily “close” to the non-analytic curve  $m_\infty(h)$  over  $\Omega$ , is to say that for every sufficiently small  $\epsilon$ -neighborhood of  $m_\infty(h)$  *in the topology*  $\tau_{\mathcal{L}^1(\Omega)}$  there exists a sufficiently large  $N$  (i.e.  $\delta$ -neighborhood of  $0 \in \mathbf{P}$ ) such that any model with  $N$  degrees of freedom (or more) must be  $\epsilon$ -close or closer to behaving like the model with non-analyticities. Hence, in at least one potentially relevant sense we do have a way of “saying how close an actually smooth curve is to a nonsmooth/nonanalytic one.” More importantly, given the right  $\Omega$ , such an  $\epsilon$ - $\delta$  deductive continuity relationship would mean that this representation is robustly  $\epsilon$ -faithful when it comes to gaining knowledge about how sufficiently large finite systems do  $\mathcal{L}^1$ -approximate the genuinely discontinuous change in magnetization of  $m_\infty(h)$ .

Batterman’s point, of course, is that there are *other* “relevant” ways (topologies) with respect to which no such continuous relationship can be established. In particular, let us define the property space  $\mathbf{Q}_{\text{Analytic?}} := \{\text{yes}, \text{no}\}$  equipped with the discrete topology  $\tau_{\text{Analytic?}}$ , and the  $\mathbf{Q}_{\text{Analytic?}}$ -property homomorphism  $\psi_{\text{Analytic?}}$ , mapping all models in  $s_N \in \mathbf{S}_{N \leq \infty}^d$  with analytic partition functions (and so analytic magnetization density curves) to the property  $\text{yes} \in \mathbf{Q}_{\text{Analytic?}}$  and the rest to the property  $\text{no} \in \mathbf{Q}_{\text{Analytic?}}$ . In contrast to the more “fine grained” topological structure of our  $\tau_{\mathcal{L}^1(\Omega)}$  the fact that  $\tau_{\text{Analytic?}}$  is the discrete topology on  $\mathbf{Q}_{\text{Analytic?}}$  means that *the only elements in  $\mathbf{S}_{N \leq \infty}^d$  that get mapped to a proper neighborhood of a non-analytic partition function must be non-analytic themselves*. Hence, the only way to get “close” to a model with a discontinuous magnetization curve like  $m_\infty(h)$  in the  $\tau_{\text{Analytic?}}$  sense is if you have  $N = \infty$  degrees of freedom. Also, the only infinite sequences of elements from  $\mathbf{S}_{N \leq \infty}^d$  converging with respect to  $\tau_{\text{Analytic?}}$  to an element that is non-analytic must have only non-analytic elements after some point in the sequence. So, since finite models in  $\mathbf{S}_{N \leq \infty}^d$  are all mapped to  $\text{yes} \in \mathbf{Q}_{\text{Analytic?}}$  no sequence of finite models can ever converge to



### 2.3. $\epsilon$ -FIDELITY AND PROXIMITY TO PATHOLOGY

a non-analytic one *in the qualitative topology*  $\tau_{Analytic?}$ . This is why Batterman says there is a qualitative difference between, for example, Ising models with  $N < \infty$  and the so-called “thermodynamic limit” with  $N = \infty$  degrees of freedom. Though a thermodynamic limit model  $s_\infty$  may count as *the limit* of some sequence  $(s_N)_{N \in \mathbb{N}}$  in the sense that the  $s_N$  converge to  $s_\infty$  with respect to *some* topology, *they cannot converge to “the” thermodynamic limit with respect to the topology*  $\tau_{Analytic?}$ . That is to say, with respect to  $\tau_{Analytic?}$ , “the thermodynamic limit” is not even a limit.

This kind of distinction has an analogy in our rationallium example above. In that case, though the models in our sequence  $((a_i, b_i))_{i \in \mathbb{N}}$  defined above came *close* (in the  $\mathbb{Q}$  space topology defined there) to having a volume ratio of 2 to 1, for no  $i < \infty$  was it the case that  $\psi((a_i, b_i)) = 2$ . There is a “qualitative distinction” between failure and success in meeting the challenge exactly. Consider the topology  $\tau_{2?} := \{\mathbb{R}, \mathbb{R} \setminus \{2\}, \{2\}, \emptyset\}$  that is also well defined on  $\mathbb{R}$ . Imagine that instead of using the property space  $\mathbb{Q}$  defined in section 2.3 which was equipped with the natural topology on the real numbers, we had selected the property space  $\mathbb{Q}_{2?}$  also consisting of the set of real numbers  $\mathbb{R}$ , but this time equipped with the topology  $\tau_{2?}$ . For this property space with this “picky” topology it is very difficult to get “close” to the property  $2 \in \mathbb{Q}_{2?}$ . Using the same  $\psi$  as in section 2.3, the only way for a model  $s \in \mathbb{S}$  to be in a proper neighborhood of 2 is if  $\psi(s) = 2$ . But as we saw above, it follows from the irrationality of  $\sqrt{2}$  that  $\psi[\mathbb{S}] \cap \{2\} = \emptyset$ . The  $\mathbb{Q}_{2?}$  property space recognizes a *qualitative difference* between models that have volume ratio of exactly 2 to 1, and those that do not, and unrelentingly rules out every  $s \in \mathbb{S}$  as not close enough.

The important point to take from this example is that even though *no sequence* of models in  $\mathbb{S}$  can converge to the property  $2 \in \mathbb{Q}_{2?}$ , there *does exist* a sequence of model that converges to the property 2 in our original  $\mathbb{Q}$ -property space. Both facts can be true simul-

## 2.4. CONCLUDING REMARKS: TOWARDS THE POINT OF PATHOLOGIES

taneously, and both facts are just another way of expressing what we learned in section 2.3: While the rationallium challenge cannot be solved *exactly*, we can find pairs that come arbitrarily close to solving the problem (in the natural real numbers topology sense of closeness). And though the former fact is problematic for perfectly faithful representations, the latter makes way for knowledge gained through an  $\epsilon$ -faithful representation.

The same thing is happening in the case of phase transitions. While no subset of finite models from  $\mathbf{S}_{N \leq \infty}^d$  can get  $\tau_{Analytic?}$ -close to a model exhibiting Kadanoff’s “pure thermodynamic phases,” it can also be true that sufficiently large but finite models in  $\mathbf{S}_{N \leq \infty}^d$  get close in other kinds of topologies such as  $\tau_{\mathcal{L}^1(\Omega)}$ . Further, topologies like  $\tau_{\mathcal{L}^1(\Omega)}$  can track physical similarity relationships in which we are interested.<sup>38</sup> The remaining question, to which we will turn in our final section, is whether or not (and if so how) these topologies can respectively enable knowledge about actual physical targets.

## 2.4 Concluding Remarks: Towards the Point of Pathologies

The example of section 2.3.1 showed us that though anchor points in the respective property spaces are essential to generating the kind of  $\delta$  to  $\epsilon$  deductions establishing  $\epsilon$ -fidelity, the presence of an anchor model in the space  $\mathbf{S}$  is not necessary even when deductive continuity can be established. Further, when an anchor model is admitted into our space of models, the fact that it can be qualitatively distinguished from non-anchor models with respect one topology imposed on a potential property space does not preclude the existence of other property spaces equipped with other topologies that may allow for  $\epsilon$ - $\delta$  continuity with

---

<sup>38</sup>Cf. the related point made by Butterfield (2011, §3.3.2).

## 2.4. CONCLUDING REMARKS: TOWARDS THE POINT OF PATHOLOGIES

respect to a certain sort of deduction. In particular, we saw that both the  $\tau_{\mathcal{L}^1(\Omega)}$  topology and the  $\tau_{Analytic?}$  topology may be used on their respective property spaces to help draw conclusions about how sufficiently large but finite systems might behave  $\tau_{\mathcal{L}^1(\Omega)}$ -similarly to a (pathologically infinite) non-analytic anchor model even though they cannot possibly come  $\tau_{Analytic?}$ -close. Hence, in contrast to what we have been calling “narrow scope” matching accounts, this kind of result means that we need not worry about our anchor models being pathological, because  $\tau_{\mathcal{L}^1(\Omega)}$ -proximity to a pathological anchor does not mean an actual physical target must also be  $\tau_{Analytic?}$ -close (and hence pathologically infinite according to statistical mechanics). Two questions remain: First, is  $\tau_{\mathcal{L}^1(\Omega)}$ -proximity (or proximity using another “fine grained” topology) to a pathological model of epistemic merit when it comes to learning about actual (finite) physical targets? And second, what role might pathological anchor models themselves play in helping us to gain knowledge of non-pathological physical targets?

As emphasized in section 1.3.4, our attempts to extract stable phenomenal patterns are ultimately indebted to the kind of theoretical presuppositions we employ. No experimentalist ever directly observes that the phenomenal magnetization relationship  $m(h)$  is either *analytic* or *non-analytic*. She observes is the data not the phenomena. Because there is an inevitable gap in moving from the data to the phenomena, we must be circumspect (though not necessarily skeptical) about our presuppositions. The epistemological situation is analogous to the simple example considered in section 2.2.2 when we discussed measuring some value  $X$  under “noisy” conditions. Expecting that a given abstraction is *exactly* captured by any particular curve is incredibly difficult to justify, even with the best equipment and unlimited time for repeated measurements. In contrast, it is possible (with sufficient techniques, tools, and repetition) to justify that an abstraction is *close* to a given curve if closeness is spelled out

## 2.4. CONCLUDING REMARKS: TOWARDS THE POINT OF PATHOLOGIES

in the right way. So, for example, by committing to the magnetization relationship  $m_\infty(h)$  *up to some  $\tau_{\mathcal{L}^1(\Omega)}$ -neighborhood*, we are able to avoid the prohibitive difficulty of justifying the claim that  $m_\infty(h)$  is the *absolutely correct* phenomenal relationship for every  $h \in \Omega$ , while still rigorously establishing that “on the whole”  $m_\infty(h)$  is not too far off over the entire interval  $\Omega$ . *It is an epistemically untenable position to insist that data gained from observing a phase transition can only ever count as evidence of a non-analytic (or an analytic) phenomenal relationship.* However, by allowing the kind of wiggle room provided by taking advantage of a fine grained topology such as  $\tau_{\mathcal{L}^1(\Omega)}$  we can avoid this kind of epistemically untenable position.

Such a strategic relaxation of the epistemic commitments to a given phenomenal pattern is especially salient given the pathologies associated with systems that are capable of exhibiting “pure phase transitions” according to statistical mechanics. Though it is impossible to observe that a phenomenal relationship is either analytic or non-analytic, we can be fairly confident that a boiling pot of water on the stove does not have an infinite number of molecules.<sup>39</sup> Hence, representing phase transition behavior for such finite systems without requiring that they meet conditions entailing that they have an infinite number of degrees of freedom is beneficial. As already stated, this ability to avoid the inference that pathological constraints imposed on a particular model must be matched by a physical target is a significant advantage.

Given such arguments against epistemically committing to exclusively non-analytic phenomenal patterns in phase transitions, we might wonder if there is any epistemological role for such pathological models to play. The answer is “yes.” Chapter 4 will be devoted (in

---

<sup>39</sup>Note, this does not necessarily entail that a system must have a finite number of degrees of freedom, but it is suggestive of the latter claim.

## 2.4. CONCLUDING REMARKS: TOWARDS THE POINT OF PATHOLOGIES

part) to explicitly demonstrating how this can be done for the case of a particular pathology that in chapter 3 we will argue occurs in Einstein’s general theory of relativity. To get a sense of how this will be done, we might note that though the existence of an anchor model is not necessary in order to gain the kind of knowledge made available by  $\epsilon$ -faithful representation, such models can serve as stark paradigms of the kind of phenomenal behavior we can expect (up to  $\epsilon$ ) of non-pathological systems. Pathological anchor models do, after all, *anchor* the kind of behavior that less pathological models can only ( $\epsilon$ ) approximate. So though pathologies preclude the possibility of these anchors being well matched to any actual physical targets, as anchors they do (in a sense) exhibit in an unadulterated way the kind of general patterns we might expect if we consider systems that should be well matched with their non-pathological  $\epsilon$ -neighbors.<sup>40</sup>

---

<sup>40</sup>As we will elaborate in chapter 4 such “general patterns” can be stated with more precision in terms of the concept of *universality phenomena*. Batterman’s admirable research on this kind of phenomena and the sort of knowledge that may be gained from trying to understand it is the reason why he argues for the importance of (also) appreciating the kind of “qualitative distinctions” in what he calls “singular limits” as referenced in section 2.3.2. See e.g. (Batterman, 2002, 2005, 2010, 2011) for further discussion.

## Chapter 3

# Proving the Principle<sup>1</sup>

In his initial formulation of the general theory of relativity, Einstein’s proposal that *freely falling gravitating massive bodies follow geodesic paths* was submitted as an independent fundamental principle. By adopting this “geodesic principle” to supply the theory’s law of motion, Einstein was immediately able to recover both the free-fall motion of bodies in non-relativistic regimes and the previously anomalous precession of the perihelion of Mercury. Over the last century numerous ostensible proofs claiming to have derived the geodesic principle from Einstein’s field equations have been developed. As a result physicists and philosophers of science alike frequently herald Einstein’s theory for having the unique distinction of being able to derive its dynamical “law of motion” from its own field equations.

In this chapter we will critically survey the multiple attempts to derive the geodesic principle in the context of Einstein’s theory. Grouping these results into three major families, which we refer to as (1) *limit operation proofs*, (2) *0<sup>th</sup>-order proofs*, and (3) *singularity proofs*, we will argue that none of these strategies successfully demonstrates the geodesic principle,

---

<sup>1</sup>In the following,  $\mathcal{M}$  will be taken to be a smooth, orientable, four-dimensional manifold, and  $(\mathcal{M}, g_{ab})$  will be referred to as a Lorentzian spacetime if  $g_{ab}$  is a smooth metric of signature  $(+, -, -, -)$  defined on  $\mathcal{M}$ . Excepting quoted material all further notational conventions follow that of (Wald, 1984).

canonically interpreted as a dynamical law that massive bodies must actually follow geodesic paths in Einstein’s theory.

Specifically, we will argue for the following three claims: First, limit operation proofs fail to demonstrate that massive bodies are ever guaranteed to follow geodesic paths. Second, on the contrary  $0^{th}$ -order proofs demonstrate that extended massive bodies generically deviate from uniformly geodesic paths. Moreover, the only potentially extended distributions of matter and energy that fail to avoid a uniform geodesic evolution are highly unstable, deviating from such motion under arbitrary perturbations of their angular momentum (or higher order moments). Third, thanks to certain mathematical theorems concerning distribution theory, alternative representations of massive bodies as unextended “point” particles must result either in precluding the possibility of coupling the particle to the spacetime metric in a way that is coherent with Einstein’s field equations or in having to excise the particle (and its would-be path) from spacetime entirely. This three pronged argument reveals that not only does the geodesic “law of motion” fail to be a deductive consequence of the field equations, but also any attempt to canonically interpret the geodesic principle in such a way requires that either the gravitating body is not massive, its existence violates Einstein’s field equations, or it does not exist within the spacetime manifold at all (let alone along a geodesic).

Hence, in the context of Einstein’s general theory, these results entail that models of massive bodies following perfectly geodesic paths fall under the pathological idealizations category discussed in chapter 2. While this is a problem for the canonical interpretation of the geodesic principle as providing a fundamental law of motion or dynamical equation, as we saw in section 2.3, such pathologies do not necessarily preclude using these models to gain the kind of knowledge provided by  $\epsilon$ -faithful representations. In chapter 4, we will

### 3.1. EINSTEIN AND THE CANONICAL ACCOUNT

directly investigate how this  $\epsilon$ -fidelity is possible in the face of the geodesic pathologies identified in this chapter. Our ultimate argument will be that such models when understood appropriately can serve a robust epistemological function in understanding the behavior of actual massive and extended bodies that *can* coherently couple to the spacetime metric in accordance with Einstein’s theory in the following two ways: (1) such models can be used in ( $\epsilon$ -)approximating the paths of specific massive bodies (as with the perihelion of Mercury confirmation) whenever they meet certain relevant “ $\delta$ -proximity” conditions on their volume and gravitational influence, and (2) using results discussed below in section 3.4 to establish an  $\epsilon$ - $\delta$  deductive continuity relationship, we demonstrate how pathologically geodesic models can play an important explanatory role as anchor models of classes of gravitating free-fall bodies exhibiting a specific kind of ( $\epsilon$ -)clustering confirmable in nature. We shall argue that this latter epistemological role recovers the geodesic principle in the form of a *universality thesis*, where the concept of universality will be analyzed in chapter 4.

## 3.1 Einstein and The Canonical Account

### 3.1.1 Geodesic Dynamics

Einstein’s adoption of the geodesic principle was originally thought to be an independent postulate establishing the dynamics of the theory. Not long after the debut of his general theory, however, numerous special-case results and plausibility arguments were developed suggesting that in fact the principle was not logically independent (given certain assumptions about free-fall bodies) from Einstein’s field equations themselves.<sup>2</sup> In the appendix

---

<sup>2</sup>Some of the earliest cited proofs and plausibility arguments include (Weyl, 1922, Eddington, 1923, Pauli, 1921, Einstein & Grommer, 1927, Mathisson, 1937, 1940). Though (Einstein & Grommer, 1927) has often been cited as the earliest result, the results by Eddington, Weyl, and Pauli clearly predate it.



### 3.1. EINSTEIN AND THE CANONICAL ACCOUNT

to the third edition of *The Meaning of Relativity* (1946), Einstein notes these developments concerning what he still refers to as “the law of motion” as follows:

In the initial formulation of the theory the law of motion for a gravitating particle was introduced as an independent fundamental assumption in addition to the field law of gravitation ... which asserts that a gravitating particle moves in a geodesic line. This constitutes a hypothetic translation of Galileo’s law of inertia to the case of the existence of ‘genuine’ gravitational fields. It has been shown that this law of motion - generalized to the case of arbitrarily large gravitating masses - can be derived from the field-equations of empty space alone. (Einstein, 1922*b*, p113)

Beyond crediting the apparent redundancy of postulating the geodesic principle as an independent assumption, note that Einstein explicitly characterizes the derivation result as pertaining not to some kind of test particle of either vanishing or arbitrarily small relative mass, but to *arbitrarily large gravitating masses*. The referenced result is no doubt that of (Einstein & Grommer, 1927) (and its successors), frequently considered a *locus classicus* of early demonstrations. As we shall see in section 3.1.3, since (Einstein & Grommer, 1927) was considered, at least by Einstein, to be a derivation of the geodesic principle, this work serves as an invaluable guide to how he expected the principle to be interpreted. In particular, it offers significant illumination into what Einstein came to believe was the content of his geodesic principle.<sup>3</sup>

---

Lesser known variations of these results were also offered in (Kopff, 1923), (von Laue, 1921), and (Becquerel, 1922), which were popular as texts on the new theory at the time (see (Havas, 1989, 1993) for further discussion of Einstein’s evident oversight in recognizing this early work). Despite his comments on the apparent redundancy of the geodesic principle (see below), Kennefick (2005) has argued that Einstein was very likely aware of the possibility of such special-case deductions prior to (Einstein & Grommer, 1927) as evidenced by possible fragments of an unadopted manuscript for (Einstein, 1922*b*). Moreover, Einstein was clearly aware of the possibility of a special-case deduction, which carries over to the general theory, from his *Entwurf* predecessor to the debut of the full theory in (Einstein, 1913) (see note 14 below).

<sup>3</sup>The idea behind what Einstein and Grommer identify as their preferred result is to squeeze the bodies into singular curves that are then excised from the spacetime entirely. At that point the source terms in the field equations of such a spacetime vanish, which is why in the long quote above Einstein notes that it can be derived from the equations for “empty space alone.” (The bodies have been fit entirely into the excised curves making them technically “outside of” the manifold and so not source terms of the field equations.) The demonstration is supposedly completed by their argument that, if we *were* to “replace” the excised curves, they *would* be geodesics of the vacuum solutions to the re-patched spacetimes (cf. (Infeld & Schild, 1949, p410)). This vacuum-cum-singularities technique was further developed by Einstein, along with Infeld and

### 3.1. EINSTEIN AND THE CANONICAL ACCOUNT

With these results it seemed that general relativity differed remarkably from other classical field theories such as classical electrodynamics or Newtonian gravitation. In (Einstein & Grommer, 1927), the authors highlight an apparent matter-field duality found in these classical field theories. Echoing this dichotomy Leopold Infeld and Alfred Schild later characterize this equation duality in classical field theories as follows:

Classical physics is dominated by a characteristic duality of field and matter. In Newton’s theory of gravitation as well as in the Maxwell-Lorentz theory of electromagnetism the physical laws fall naturally into two independent classes. The first class consists of the partial differential equations which (with suitable boundary conditions at infinity) determine the field in terms of the distribution and motion of the matter which “generates” it. The second class consists of the dynamical equations governing the motion of matter under the forces “exerted” by the field. (Infeld & Schild, 1949, pp408-9)

They then proceed to explain how the equations of general relativity (viz Einstein’s field equations plus the geodesic equation) fit into this picture, observing that just as in cases like classical electrodynamics, where there are two sets of equations, one set for how the fields couple with source charges (Maxwell’s equations), and another for dynamics of how “passive” charged bodies behave in those fields (the Lorentz force law), so too is there a duality corresponding to the two sets of equations in Einstein’s theory. His field equations govern how the field couples with the gravitational sources, while the geodesic equation provides the “law” for how gravitating bodies then surf the resulting metric field. In contrast to other classical field theories, however, for Einstein’s theory it now seemed that the field equations for coupling the metric to energy-momentum sources also entailed the geodesic equation for how free-fall massive bodies behave in a given geometric field. Unlike with electrodynamics and Newtonian gravitation, the dynamical equations appeared not to be

---

Hoffman, in (Einstein *et al.*, 1938, Einstein & Infeld, 1940, 1949) as Einstein became increasingly opposed to representations of matter by means of continuous fields (see section 3.1.3).

### 3.1. EINSTEIN AND THE CANONICAL ACCOUNT

logically independent.<sup>4</sup>

This “duality of field and matter” and the dichotomy of their corresponding equations, endorsed by Einstein himself, is significant for two reasons. First, it emphasizes the apparent boon for Einstein’s theory: general relativity is “special” in comparison with other classical theories because the dynamical equations of the theory appear to logically follow from its own field equations.<sup>5</sup> Second, in order to even claim that general relativity has such a special status among classical field theories, one must subscribe to a key presumption about the role of the geodesic principle active in the early decades of the theory (and still endorsed frequently today), namely, that (analogous to the role of the Lorentz force law in electrodynamics) *the geodesic principle plays the role of providing the dynamics of material bodies in the general theory of relativity*. In the following, we shall refer to this account of the role of the geodesic principle as providing the dynamics of general relativity as the *canonical account*.

In his early comments on the geodesic principle, Einstein frequently endorses this canonical account. In (Einstein, 1916) as well as his (Einstein, 1922b) lectures on the theory, Einstein refers to the geodesic equation or the principle as the “equation of motion” or “law of motion” over a dozen times, characterizing them in this way not only for application

---

<sup>4</sup>By ‘logical dependence’ here we mean derivability, perhaps under certain conditions characterizing the body in question. Of course dissolving the conceptually suspect bifurcation of bodies into “background” charged sources, which determine the field, and “passive” charged bodies that then react to the field (without generating self-forces) in this caricature of electrodynamic evolution leads to well known significant complications that have (even after over a century of effort) yet to be fully resolved (for an historical presentation and philosophical discussion of this problem see e.g. (Frisch, 2005)). As we will see, similar complications involving self-force-like effects are relevant in determining the actual motion of free-fall bodies in general relativity. The independence resulting from such a bifurcation of bodies into background sources and passive test bodies is, nonetheless, a separate notion from the *logical* independence of the dynamical equations of motion from the field equations, which (at least according to the interpretation we are now considering) exists in the electrodynamics case but not in the relativistic case.

<sup>5</sup>This distinction has been highlighted by philosophers such as Brown (2005, pp140-1) as well as the physicists who worked on this problem in the early decades (e.g. (Einstein & Grommer, 1927, Infeld & Schild, 1949)). Unlike these physicists, however, Brown astutely notes what he describes as the “limited validity” of deductions establishing exact geodesic motion, a point that we will investigate in detail below.

### 3.1. EINSTEIN AND THE CANONICAL ACCOUNT

to “particles,” but also in describing “planetary motion” (most importantly the motion of Mercury) and the motion of a “gravitating body” in general. As already indicated by the above long quote, Einstein continued to view the geodesic principle as providing the “law of motion” not only for massless test particles but also for “arbitrarily large masses” well after the theory’s initial introduction.

During this period, the canonical view was likewise frequently articulated by Einstein’s colleagues. It takes only a brief survey of the literature from the first half of the 20<sup>th</sup> century to reveal the widespread general adoption of the canonical view, with most authors taking it for granted that the geodesic principle provided the dynamics of the theory regardless of its logical independence from the field equations.<sup>6</sup> Expressions of this view were unmitigated (and sometimes even highlighted) by the apparent redundancy of postulating the geodesic principle as an independent assumption. This attitude is typified by the commentary of physicists such as Lanczos, for instance, who punctuates his demonstration by noting that his penultimate equation “is equivalent to the ‘law of the geodesic line’ which has always been considered *the natural dynamical law of general relativity*” (Lanczos, 1941, p818 emphasis added). Moreover, this canonical view of the geodesic principle as providing the dynamics is frequently cited in text books on the subject both classical (e.g. (Bergmann, 1942, pp224-5)) and contemporary (e.g. (Hobson *et al.*, 2006, pp188-90) and even (Misner *et al.*, 1973, pp475-80)).<sup>7</sup> Though our focus will be on Einstein’s interpretation of the geodesic principle and its

---

<sup>6</sup>E.g. references to the geodesic principle as providing the dynamics or law of motion in some form or another are evident in (Eisenhart, 1928, Eddington, 1923, Tolman, 1930, Dirac, 1938, Lanczos, 1941, Infeld & Schild, 1949). Some authors judiciously express the view in restricted form only as pertaining to “relatively small” masses or simply to “mass points” (e.g. (Weyl, 1922, p256) or (Bergmann, 1942, pp224-5)). As we will see, in the former case, there was still little real justification for such heuristic winks at sufficient smallness (see sections 3.1.2 and 3.4.2 below), whereas in the latter case significant difficulties abound when it comes to representing the massive point particle that supposedly follows a geodesic within the theory (see section 3.2).

<sup>7</sup>In the case of references found in contemporary texts there should be no doubt that the authors are well aware of gravitational multipole and “self force” effects resulting in non-geodesic motion (see section 3.3 below). (Misner *et al.* (1973, p479) are notably circumspect about some of these failures and later offer

### 3.1. *EINSTEIN AND THE CANONICAL ACCOUNT*

role in providing the dynamics of the theory, he was not alone in this attitude well into the mid 20th century. It is thanks to this combination of endorsements that the account of the geodesic principle providing the dynamics of Einstein’s theory is plausibly characterized as “canonical.”

#### 3.1.2 **Whither Test Particles?**

If, according to what we have described as the canonical account, the geodesic principle provides the dynamics of the general theory of relativity, we must figure out to what exactly such a dynamical principle is supposed to pertain. Who follows geodesics? A natural answer might be something like “test bodies,” the theoretical tool in the physicist tool box used to describe how certain “sources” react to the field without having to attend to the actual effects on the field values caused by the presence of the bodies in question. In the case of relativity theory, we might then answer that “it is test bodies who follow geodesics.”

While we will ultimately see that under a non-canonical interpretation something like this answer might be endorsed (section 3.5), in the following survey of geodesic demonstrations it will be of central importance to observe exactly why and in what manner ignoring the source effects of “test bodies” can be justified. That is to say, we will need to pay special attention (i) if a gravitating object is treated as a “test body” because its source effects are simply left unaccounted for, or (ii) if the object is treated as a “test body” because its source effects can be shown to be negligible (but non-vanishing) for the relevant purposes of the deduction. The hazards of leaving test body approximations unjustified (i.e. case (i)) become most vivid when we consider proofs of the geodesic principle. In cases where the field

---

explicit instruction on calculating spin effects.) Hence, such references should be taken only as evidence of the pervasive popular endorsement of the canonical view and the fact that the effect of the view’s initial adoption still lingers in contemporary conceptions of Einstein’s theory.

### 3.1. EINSTEIN AND THE CANONICAL ACCOUNT

equations and dynamical equations are starkly separated, physicists have the luxury of an apparent distinction: bodies whose source behavior is “turned on” are governed by the field equations, whereas the behavior of test bodies can seemingly be restricted to the purview of the dynamical equations alone. However, if one attempts to *deduce* the dynamical laws from the field equations, this specious luxury evaporates. We are forced in the course of the proof to simultaneously discuss the matter-energy of the field equations as the matter-energy that we ultimately hope to show obeys the dynamical equations. Hence, it is not even an apparent option to treat test matter-energy as being *entirely free* of the field equations as might happen in case (i).

As we will see, under the canonical interpretation, ignoring source effects of a body (even when they are small) can often have significant impact on the general validity of the deductions. With his characteristically sardonic wit when discussing this subject, Jürgen Ehlers, in collaboration with Ekkart Rudolph, emphasizes this challenge as follows:

The *test body approximation* is usually defined by the requirement that the contribution of the body to the metric  $g_{\alpha\beta}$  be negligible. The justification of this drastic simplification in any particular case is by no means trivial and is therefore rarely considered. Since, according to Einstein’s (and similar) field equations, the curvature within a body is of the order of the density, the “self curvature” usually dominates or is at least comparable to the “incident” or “external” curvature (even for a small iron ball near the Earth’s surface), and then it is wrong to take the metric within the body to be nearly equal to the “given,” external one in the local mechanical law  $[T^{\alpha\beta}_{;\alpha} = 0]$ . (Ehlers & Rudolph, 1977, p208)<sup>8</sup>

In case (ii) above the physicist will be able to explain why the effects of the test bodies are inconsequential in a relevant and rigorous sense and may hence be justifiably ignored. While the majority of attempts at geodesic demonstrations (certainly, at least, at the time of this quote) seemed to fall under case (i), Ehlers and Rudolph here explain that a supplementary

---

<sup>8</sup>Ehlers and Rudolph go on to explain in a parenthetical that “For this reason the mathematically elegant argument given in (Geroch & Jang, 1975) is physically not very enlightening, in our opinion.” We will return to this point in section 3.4.1.

### 3.1. EINSTEIN AND THE CANONICAL ACCOUNT

justification of these test body approximations with heuristic winks at relative smallness will not typically suffice; far more work is left unfulfilled.

If the geodesic principle is to provide a dynamics that can be legitimately used to predict the paths of actual bodies, we must find a way to draw suitable inferences from how these test particles are supposed to behave to how actual bodies behave. Unfortunately, as is well known from the case of classical electrodynamics, paying attention to the actual field-creating abilities of our (in the electrodynamics case, charged) test bodies, things become increasingly messy. Shrinking the body down to “infinitesimal” volume results in a singular charge density, and extending the particle still results in having to grapple with non-analytically expressible expansions of the effects that the particle’s own field has on its motion.

In the case of general relativity things are even more treacherous. As we shall see, not only are there self-force and spin effects to be grappled with, but also, in the case of general relativity, the presence of matter-energy, whose powers as a field source have not been artificially “turned off,” will affect the very metric that determines what counts as a geodesic. An infinite matter-energy density in general relativity is not just an aesthetically disheartening anomaly in our representation, it often results in our inability to coherently speak about the spacetime path where the singularity occurs. But if the metric becomes undefined wherever the source particle is located (if it can even be said to have a location), how are we supposed to say that it is following a “geodesic” of that metric? On the other hand, the modeling of extended bodies in general relativity leaves a good deal more freedom available for how the body’s matter-energy is distributed, making it difficult to speak generally about representations of the bodies (especially that they universally follow geodesics). As we shall see, some of these issues had already become manifest by the time of ([Einstein & Grommer, 1927](#)).

### 3.1.3 Einstein and Grommer’s “Three Ways” to get it Straight

In the introduction to their paper Einstein and Grommer lay out the same dichotomy in Newton’s theory of gravitation and classical electrodynamics between field equations and dynamical equations articulated decades later in the long quote discussed above from Infeld and Schild. Characterizing such matter-field “dualism” as “disturbing to any systematic spirit,” they proceed to identify three “ways” [Betrachtungsweisen] in the general theory of dealing with such duality (Einstein & Grommer, 1927, p3).<sup>9</sup>

Their “first way” is modeled after Newtonian gravitation, in which the field equations and the geodesic equation are posited independently.<sup>10</sup> This approach is most similar to Einstein’s initial introduction of the theory in that the field equations and the geodesic equation are postulated independently. Unlike his initial introduction of the theory, however, in this method the field equations in question are not Einstein’s full field equations:

$$G_{ab} = T_{ab} \tag{3.1.1}$$

where the *Einstein curvature tensor* on the left hand side is defined by

$$G_{ab} := R_{ab} - \frac{1}{2}g_{ab}R \tag{3.1.2}$$

and the right hand side of the equation represents the flow of matter-energy from any per-

---

<sup>9</sup>Unless noted otherwise, this and all below translations of (Einstein & Grommer, 1927) are thanks to the gracious assistance of Bihui Li.

<sup>10</sup>This method can be thought of as being modeled after the field theoretic accounts of Newtonian gravitation in the sense that there too background sources might be represented by singular points generating a gravitational potential field  $\phi$ , where  $\phi$  is a solution to Laplace’s equation (i.e. Poisson’s equation with  $\rho = 0$ ) with suitable boundary conditions at the singular points. This potential field is then surfed by test bodies satisfying the equation of motion  $-\nabla^i \phi = \frac{d^2 x^i}{dt^2}$ , which of course comes from Newton’s laws of motion and gravitation. Einstein emphasizes this analogy explicitly in (Einstein & Infeld, 1949, pp209-10), his final collaboration with Infeld on the subject.



### 3.1. EINSTEIN AND THE CANONICAL ACCOUNT

spective. Instead, for their “first way” they specify that the relevant field equations are the *vacuum* field equations where equation (3.1.1) reduces to the equation:

$$R_{ab} = 0 \tag{3.1.3}$$

once the energy-momentum tensor field is made to vanish everywhere. At first blush, the appeal to the vacuum field equations in this account may strike the reader as somewhat backwards. According to the dichotomy discussed above, it would seem that in this method Einstein and Grommer are dealing with the uncomfortable distinction between “test energy-momentum” and the energy-momentum sources contributing to the gravitational field by eliminating the sources while keeping the test bodies. But in the discussion above, it was the energy-momentum *test bodies* facing conceptual complications, not the “background” sources. So in their “first way” it might appear that they are getting things the wrong way round, having eliminated the source energy-momentum while retaining only the conceptually suspect test bodies.

One way of seeing why they specify the vacuum equations in this case is to consider why (from the perspective of 1927) Einstein might have wished to employ the two sets of equations to generate predictions.<sup>11</sup> In particular, we might consider how he would have calculated the perihelion of Mercury at that time. Once the Schwarzschild solutions had been discovered, it was possible to determine the stable geodesics of the metric and from there calculate the perihelion of the orbit. So modeling the sun (minus all the planets) with the Schwarzschild solution, we could apply this calculation schema to the case of a Mercurial test body in such a background metric. But observe, the Schwarzschild solution is a *vacuum*

---

<sup>11</sup>See Einstein’s comments on this strategy in (Einstein, 1995, p310).

### 3.1. EINSTEIN AND THE CANONICAL ACCOUNT

solution. According to our application of this model, all the matter and energy of the Sun is to be found not in the spacetime manifold but “in” the singularity of the Schwarzschild Solution “located” at the origin of the coordinate system. Einstein and Grommer could consider the first way to be a possibility even though there is literally no place for energy-momentum sources in the manifold, because hiding the sources in the singularity works so well in this kind of application. Einstein and Grommer’s selection of the vacuum equations in the “first way” is indicative of a significant shift in how Einstein in particular began to prefer to represent matter-energy in his (as he saw it, not yet complete) theory. This attitude becomes even more apparent in their response to the next method.

The problem with the “first way” of course is that rather than dissolve the aforementioned discomfort with matter-field dualities when it comes to the general theory, it exacerbates the dichotomy. In contrast, according to their “second way,” *all* matter-energy is represented via a continuous and singularity-free energy-momentum tensor field  $T_{ab}$ . Unlike their “first way” method, this time they seem to get things the right way round when it comes to the elimination of the potentially suspect test bodies. They keep only the source matter-energy of the tensor  $T_{ab}$  while eliminating all appeals to test matter-energy.

After noting that as a consequence of (3.1.1), the total divergence of the energy-momentum tensor vanishes, without any calculation or further explanation, they make the following claim:

If one assumes that matter is arranged along narrow “world-tubes” one obtains from this by an elementary consideration the theorem that the axes of those “world-tubes” are geodesic lines (in the absence of electromagnetic fields). This means: the law of motion is a consequence of the field law.<sup>12</sup>

It is difficult to speculate which “elementary consideration” establishes their demonstration.

---

<sup>12</sup>Translation of quote from (Havas, 1989, p240).

### 3.1. EINSTEIN AND THE CANONICAL ACCOUNT

Though by 1927 special-case derivations of the geodesic principle from Einstein’s non-vacuum field equations had gained substantial proliferation, in the intervening decade since the debut of his general theory, Einstein never published any discussion or recognition of such (apparent) redundancy. Given his well known reputation for neglecting the literature, it is possible (though remarkable in light of his familiarity with a number of the authors)<sup>13</sup> that Einstein was not even aware of the abundance of such results.<sup>14</sup> Eddington’s plausibility result in particular would appear to be paradigmatic of Einstein and Grommer’s “second way” approach, but it requires significant symmetry assumptions about the world-tubes in question in order to establish geodesic motion. Whether they were aware of these earlier results or simply referring to their own margin calculations, for reasons that will become evident in section 3.3, it is difficult to imagine that Einstein and Grommer’s unexplicated

---

<sup>13</sup>See (Havas, 1989) for detailed discussions on this point.

<sup>14</sup>Einstein’s decade of silence (at least in publications) on the derivability should not be taken as evidence of his ignorance of special-case derivations. It has been recently argued by Kennefick (2005) that there is evidence that he was quite familiar with the possibility of special-case results. In particular the geodesic motion of pressureless dust matter, which transfers to the full theory, was derived within the *Entwurf* theory. The easily transferred *Entwurf* result in question can be understood by considering the following elementary derivation: Suppose matter takes the form of a “pressureless dust” such that the energy-momentum tensor field can be written

$$T^{ab} = \rho U^a U^b$$

where the  $U^a$  have been normalized to be unit timelike. Then, if the covariant derivative of the left hand side vanishes we have

$$0 = \nabla_a(\rho U^a)U^b + \rho U^a \nabla_a U^b$$

but contracting with  $U_b$  annihilates the second term leaving us with

$$0 = \nabla_a(\rho U^a).$$

So plugging this back into the second equation, at spacetime events where  $\rho \neq 0$  we can divide through by  $\rho$  giving us that the “dust matter” there obeys the geodesic equation

$$0 = U^a \nabla_a U^b.$$

In a recently uncovered fragment of notes evidently intended for his (Einstein, 1922b), Einstein claims that his field equations “already contains [sic.] the divergence equation and with it the laws of motion of material points,” suggesting that he remained aware of this kind of result during the intervening decade. It should go without saying that success in such a pressureless dust derivation does not generalize to arbitrary applications of the principle (nor is its application in certain cosmological models above reproach).

### 3.1. EINSTEIN AND THE CANONICAL ACCOUNT

“second way” derivations could have been terribly general, despite their tone to the contrary.

In any case, they immediately abandon this victory over the matter-field dualism in general relativity, rejecting such “second way” derivations on the grounds that the use of a continuous energy-momentum tensor field  $T_{ab}$  to represent the distribution of matter-energy throughout the manifold is suspect:

It looks as though the general theory of relativity has already overcome that annoying dualism. This would be the case if we had already arrived at a representation of matter through continuous fields, or if we were at least convinced that one day we will arrive at it. But there can be no question of that happening. All attempts in the last years to explain the elementary particles of matter through continuous fields are failures. The suspicion that this is ultimately not the correct route to understanding material particles has become very strong in us.

This suspicion of the energy-momentum tensor was by no means a sudden development in Einstein’s attitude. Such comments echo cautions voiced by Einstein from the very beginning of his presentation of the general theory. He expresses wariness about such a representation of matter-energy, for instance, in his ([Einstein, 1922b](#)) lectures as follows:

In reality, matter consists of electrically charged particles, and is to be regarded itself as a part, in fact, the principal part, of the electromagnetic field. It is only the circumstance that we have no sufficient knowledge of the electromagnetic field of concentrated charges that compels us, provisionally, to leave undetermined, in presenting the theory, the true form of this tensor. From this point of view it is at present appropriate to introduce a tensor,  $T_{\mu\nu}$ , of the second rank of as yet unknown structure, which provisionally combines the energy density of the electromagnetic field and that of ponderable matter; we shall denote this in the following as the ‘energy tensor of matter’. ([Einstein, 1922b](#), p85)<sup>15</sup>

And in his perhaps most poetic (and well known) rejection of such a continuous energy-momentum tensor field representation, in 1936 Einstein offers the following illustration of this attitude:

[General Relativity] is sufficient - as far as we know - for the representation of the observed facts of celestial mechanics. But it is similar to a building, one wing of which

---

<sup>15</sup>See also his reflection on these hesitations in ([Einstein & Rosen, 1935](#), note 3)

### 3.1. EINSTEIN AND THE CANONICAL ACCOUNT

is made of fine marble (left part of the equation [(3.1.1)]), but the other wing of which is built of low-grade wood (right side of equation [(3.1.1)]). The phenomenological representation of matter is, in fact, only a crude substitute for a representation which would do justice to all known properties of matter. (Einstein, 1995, p311)

Einstein’s disparagement of the energy-momentum tensor as analogous to “low-grade wood” has to do with its representation of matter-energy by means of the continuous tensor field. He describes it as a “phenomenological representation” because such a continuum representation is so close to the representation of matter in continuum mechanics as taking the form of a continuous medium (as is phenomenologically apparent) rather than an atomistic or quantum form.<sup>16</sup> Einstein’s resistance to the “low-grade wood” representation of matter-energy particularly in the context of determining the motion of bodies was intimately tied to his hopes for a unified theory, and the final remark of (Einstein & Grommer, 1927) explicitly speculates about their preferred “third way” methods leaving room for integration with the “quantum theory of matter.”<sup>17</sup> Though such hopes failed to come to fruition, save for a brief wavering in 1935,<sup>18</sup> Einstein would continue to resist the representation of matter-energy by means of a continuous tensor field in favor of a singularity approach for the remainder of his life, frequently voicing his skepticism of “low-grade wood” approaches.<sup>19</sup>

Einstein and Grommer’s “third way” avoids both the “low-grade wood” representation of background sources with a continuous tensor field as well as suspect appeals to test bodies. Instead (in the absence of electromagnetism), it makes use of the vacuum field equations alone, attempting to hide all matter-energy along singular “world-lines” of the manifold. In

---

<sup>16</sup>Cf. (Einstein, 1922b, pp52-3)

<sup>17</sup>His early hopes (later dashed) that attending to the motion of bodies may yield insight into such unification have been recorded by collaborators such as Infeld (1980). For discussions see (Pais, 2005, Howard, 1990, Earman & Eisenstaedt, 1999).

<sup>18</sup>This wavering was in response to a persistent challenge posed in his correspondence with Ludwik Silberstein (see (Havas, 1993) for a detailed review of this controversy) and only lasted for a period of months surrounding his publication of (Einstein & Rosen, 1935).

<sup>19</sup>See e.g. (Einstein *et al.*, 1938, Einstein & Infeld, 1940, 1949)

### 3.1. EINSTEIN AND THE CANONICAL ACCOUNT

the conclusion they characterize their result as a (special-case) demonstration that these “singular world-lines” obey the geodesic principle, stating that “[i]f one understands masses in the gravitational field as singularities, then the law of motion is fully determined by the field equations.”

Einstein’s ultimately preferred singularity approach to the representation of matter-energy significantly influenced his interpretation of the geodesic principle. His adoption (and somewhat mistaken interpretation) of the singularity method makes it clear why in his 1946 appendix to (Einstein, 1922b) he thought he could characterize the geodesic “law” as applying not just to test matter-energy but to “arbitrarily large bodies.” By making use of the singularity results, he believed he was free to hide as much matter-energy as he likes in the singular “world-lines,” while still (ostensibly) being able to derive the geodicity of such curves. In section 3.2 we will critically review the incoherence of such “third way” strategies, particularly in attempting to show that such “world-lines” are geodesics.<sup>20</sup> But for now it is worth noting that though their introduction of the singularity method is initially characterized as a representation of elementary particles, Einstein quickly shifts the auspice of his “derived” principle to include large composite bodies such as Mercury as well.<sup>21</sup> Einstein’s dynamical interpretation of the principle did not hinge on the ability to treat bodies obeying the principle as arbitrarily small, nor did he see the proper interpretation of the dynamical role of the principle as subject to the uncomfortable matter-field duality found in classical electrodynamics and Newtonian gravitation. In fact, his work with Grommer strongly

---

<sup>20</sup>To avoid a tempting conflation, note that Einstein and Grommer’s “three ways” are distinct from what we will below (in sections 3.2, 3.3, and 3.4) classify as the *three general families of deductions*. Though the method of “singularity proof” of section 3.2.1 uses Einstein and Grommer’s “third way” strategy, both the 0<sup>th</sup>-order proofs and limit operation proofs of sections 3.3 and 3.4 clearly count as “second way” strategies according to the Einstein and Grommer classification, breaking any compelling analogy.

<sup>21</sup>Moreover, early post-Newtonian confirmations of the two-body motion of stellar objects is often credited back to the work in (Einstein *et al.*, 1938) and its successors, which likewise adopts the singularity method.

### 3.2. SINGULARITY PROOFS

indicates that by 1927 he viewed the evident derivability of the principle by means of the singularity method as a significant triumph in this respect. In his view, by making use of such singularity methods, Einstein could allow the geodesic principle to play a dynamical role *for actually massive bodies* without (any longer) having to succumb to such dualism. Unfortunately, singularity proofs, both those using Einstein’s methods as well as those using more sophisticated methods, ultimately fail to establish the geodesic principle in a way that is compatible with his field equations.

## 3.2 Singularity Proofs

The family of proofs which we will refer to as *singularity proofs* really consist of two distinct subclasses. The first subclass follows Einstein and Grommer’s original “third way” method in which they attempt to use true singularities in the manifold in order to represent matter-energy. These singularities in the manifold are then (somehow) supposed to be shown to be geodesic. With the mathematical advances in distribution theory, these true singularity proofs were succeeded by the second subclass, which attempts to leave the metric well defined at the location of the geodesic following particles by coupling it to energy-momentum tensor distributions. In the next two subsections, we will consider each of these in turn.

### 3.2.1 The Geodesic that Wasn’t There

As already hinted, the most perspicuous difficulty with Einstein’s method of deducing the geodesic principle for particles represented as singularities in vacuum solutions is that (strictly speaking) the supposed path of such geodesic following particles is not even in the spacetime manifold. In (1995, p12) Earman poetically summarizes this “perplexing” strategy

### 3.2. SINGULARITY PROOFS

with the explanation that “to speak of singularities in  $g_{ab}$  as geodesics of the spacetime is to speak in oxymorons.” The proponent of such a “vacuum-cum-singularity” technique is faced with the rather paradoxical challenge of explaining in what sense we can say that a singular curve (ostensibly constituted by the *missing* points in the manifold) is actually a geodesic of the spacetime from which it is absent. Not only is no metric defined at the singularity, but technically there are not even any spacetime events there: The “geodesic” doesn’t exist. By eliminating the “low-grade wood” representation of matter-energy sources, Einstein dodged the difficulties associated with using continuous representations of energy-momentum that might restrict the generality of the principle (see section 3.3 below) but only at the cost of having to justify the geodicity of a metric-less hole in spacetime.

Though Einstein and Grommer avoid elaborate consideration of this challenge, their strategy might (briefly) be characterized as follows: splitting the first order perturbations of the Minkowski metric in the neighborhood of the singularity into an “exterior” ( $\gamma_{\alpha\beta}^{(ext)}$ ) part resulting from sources “far” from the singularity and an “interior” ( $\gamma_{\alpha\beta}^{(int)}$ ) part resulting from the ostensible presence of the body “at” the singularity, they then argue that in their chosen coordinate system,  $\gamma_{\alpha\beta}^{(ext)}$  obeys the constraint that

$$\frac{\partial \gamma_{44}^{(ext)}}{\partial x_\mu} = 0$$

along coordinates of the  $x_4$ -axis where they locate the singularity.<sup>22</sup> The suggested implication then is that for a *second* singularity-free spacetime, whose metric is given by a  $\gamma_{\alpha\beta}^{(ext)}$  correction to the Minkowski metric, the  $x_4$ -axis (not in the domain of the first metric) is

---

<sup>22</sup>Of course, technically the metric is not well defined at those coordinates, but they claim to avoid this problem by stipulating that, since  $\gamma_{\alpha\beta}^{(ext)}$  is generated by “external” sources, it should be regular in the neighborhood of the singularity. The (suspect) intimation being that for this reason it can be unproblematically extended across the coordinates of the singularity.



### 3.2. SINGULARITY PROOFS

a geodesic. They seemingly take for granted, however, that we must associate the singular boundary in the former spacetime with the corresponding “filled in” points along the  $x_4$ -axis in the latter spacetime, perhaps due to the appearance of an embedding relationship suggested by the similarity in the coordinatizations of the respective spacetimes. Unfortunately, the claim that Einstein and Grommer’s second singularity-free spacetime can tell us something about the nature of the singularity of the original spacetime is spurious. A similarity in the coordinates used to refer to the singularity in one spacetime and the coordinates of a second spacetime without a singularity at those coordinates is not enough to infer that the second spacetime is a “filling in of the singularity.”

In an attempt to vindicate the vacuum-cum-singularity strategy, Infeld and Schild concede that “[c]learly, the statement that a singular line is (or is not) a geodesic has no meaning” (Infeld & Schild, 1949, p410).<sup>23</sup> They proceed to argue that the geodesic principle might nonetheless be proven by means of the vacuum-cum-singularity strategy, if it is once again asserted that the principle is (at least) germane for a certain kind of representation of test particles:

Physically, we can consider a sequence of particles, with masses tending to zero, and a corresponding sequence of gravitational fields. In the limit  $m = 0$  we obtain a limiting world line along which the limiting gravitational field, the background field, is continuous. We must think of the background field as being assigned *a priori*; the geodesic “postulate” refers to the limiting world line in this continuous field and is thus meaningful.

Recall, Einstein claimed (as late as 1946) that the vacuum-cum-singularity method can be used to derive the postulate for “arbitrarily large masses.” The move of restricting their geodesic result only to this specific variety of test particles, which we will refer to as *Infeld-*

---

<sup>23</sup>Infeld, Einstein’s long time collaborator on the motion of bodies, became one of the principle champions of singularity methods (both the vacuum-cum-singularity method and then later the distributional method) well after Einstein’s final contributions to the problem (Einstein *et al.*, 1938, Einstein & Infeld, 1940, 1949, Infeld & Schild, 1949, Infeld, 1954, 1957, Infeld & Plebanski, 1960).

### 3.2. SINGULARITY PROOFS

*Schild* or *IS-particles*, constitutes a strategic retreat from Einstein’s position. Infeld and Schild’s derivation might hence be thought of as an attempt to embrace case (ii) considered in section 3.1.2 by trying to justify why the effects of the “test” body can be ignored. By restricting their results to these IS-particle sequences of spacetimes, Infeld and Schild were forced to limit their result to particles of *arbitrarily small* mass but now had the chance of explaining why we can associate the geodicity of a curve  $\gamma$  of the “background field” with the character of the singularities in spacetimes with  $m \neq 0$ : The background spacetime is a “limit” of the singular spacetimes, and the coordinates in each of these spacetimes demarcating the singularity are the same as those used to locate  $\gamma$  in the background spacetime.<sup>24</sup>

The limiting procedure Infeld and Schild use is fatally flawed. Though there is a coordinate similarity of the “limit” spacetime and the singular spacetimes (speciously) suggesting an embedding relation, the singularity will exist for every one of the  $m \neq 0$  spacetimes in the run up to the supposed “limit.” For every spacetime short of the background one, the “behavior at  $\gamma$ ” will remain undefined, obscuring the sense in which the singular spacetimes are “approaching” the background one. Again, there is no rigorous sense in which the singular behavior of the sequence of spacetimes converges to a non-singular background spacetime, making references to “*the* limiting behavior” literally nonsensical.<sup>25</sup>

Infeld and Schild’s attempts to derive the geodicity of singularities in the manifold by considering perturbations in the boundary conditions that could be taken to indicate the presence of arbitrarily small matter-energy located at the singularity ultimately failed. Though it is possible to use surface integral techniques, integrating around the singularity to suggest that there is (something like) matter-energy “hidden” so to speak at the undefined (singular)

---

<sup>24</sup>Actually, unlike Einstein and Grommer, they attempt to use (proto-)geodesic completion methods to covariantly specify the singular “points.” Such completion methods unfortunately remain insufficient.

<sup>25</sup>There may not even be a pathology-free (e.g. Hausdorff) way to fill in the singularity (see (Geroch *et al.*, 1982)). See (Torretti, 1996, pp178-9) for further discussion of this fallacy.

### 3.2. SINGULARITY PROOFS

boundary region, there is no way of rigorously discussing what goes on “at the singularity” of such vacuum solutions, and in particular, no way of inferring geodicity. Einstein and Grommer’s “third way” vacuum method hence turns out to be unsalvageable even with a retreat to the arbitrarily small IS-particles.

By 1954, even Infeld had turned to a kind of compromise between the “second way” appeal to a non-vanishing energy-momentum tensor field and the “third way” attempts to concentrate matter-energy onto a world-line where the metric diverges (Infeld, 1954, 1957). In this method Einstein’s “low grade wood” is replaced (metaphorically speaking) by a kind of sturdier (but ultimately poisonous) “pressure treated wood” through the introduction of energy-momentum tensor distributions.

#### 3.2.2 Distributional Energy-momentum

The idea behind distribution proofs of the geodesic principle is to concentrate all the matter-energy of a (would-be) geodesic following particle onto a one-dimensional (often timelike) curve  $\gamma$ . Once this is done, the task is to deduce from Einstein’s field equations (or a generalization of them) that  $\gamma$  must be a geodesic. In contrast to the singularity proofs of the last section, proofs using distributional energy-momentum do not use the vacuum field equations (3.1.3) ultimately preferred by Einstein. Instead a non-vanishing energy-momentum tensor distribution on the right side of the equation is used to represent the particle. In a sense then, distribution proofs are similar to Einstein and Grommer’s so-called “second way” demonstrations in that, like those proofs, they appeal to field equations coupling the geometry of the manifold (via the Einstein curvature) to non-vanishing energy-momentum sources.

Unlike “second way” demonstrations, however, distribution proofs do not (strictly speak-

### 3.2. SINGULARITY PROOFS

ing) make use of Einstein’s original field equations (3.1.1). In “second way” demonstrations, the objects equated ( $G_{ab}$  and  $T_{ab}$ ) are typically supposed to be smooth tensor fields defined on the manifold. In distributional (singularity) proofs, on the other hand, the energy-momentum field on the right hand side of the equation is a *distributional* object in some neighborhood of  $\gamma$ : assuming no electromagnetic contributions in that region, its support is restricted to a measure-zero region (i.e. the one-dimensional region  $\gamma$ ). Yet in order to attribute non-vanishing matter-energy to the “particle,” we do not want integrals of the field in that region to vanish. Hence, such proofs make use of an energy-momentum *tensor distribution* that is defined by its action on a space of well behaved (mathematical) “test” objects.<sup>26</sup> The definition of the space of tensor distributions offers a natural extension of the space of locally integrable tensor fields in a way suggesting that we can “integrate” certain distributions concentrated on measure-zero regions without the integral necessarily vanishing (details are reviewed in appendix A).

In order to conduct a distributional derivation of the geodesic principle from “the” field equations of general relativity, Einstein’s original equations (3.1.1) must (at least implicitly) be modestly generalized by saying:

$$G_{ab} = T_{ab} \quad \text{as tensor distributions} \tag{3.2.1}$$

I.e. if  $\mathcal{M}$  is an orientable manifold, for all  $\phi^{ab} \in \mathcal{T}_0^2(\mathcal{M})$  with compact support:

---

<sup>26</sup>The use of the term ‘test’ in the context of distribution theory is only incidentally similar to the use of the term to refer to “test bodies.” The former are well behaved sets of mathematical objects on which distributions act, the latter (as already discussed) refers to a kind of theoretical representation signifying entities that react to physical fields but do not act as (significant) sources of those fields. Hence, there is no room for the two usages to be equivocated despite the unfortunately abundant opportunity for confusion.

### 3.2. SINGULARITY PROOFS

$$\int_{\mathcal{M}} G_{ab} \phi^{ab} vol = \int_{\mathcal{M}} T_{ab} \phi^{ab} vol \quad (3.2.1a)$$

where  $vol$  can be any volume element defined on  $\mathcal{M}$ .<sup>27</sup> Distributional proofs hence constitute a kind of compromise between Einstein and Grommer’s “second” and “third” way methods of derivation. Equation (3.2.1) is not a vacuum equation, but rather than taking a “low grade wood” approach to representing the matter-energy of the particle as a continuously distributed field, it (in a sense) allows us to specify the “particle’s” four-dimensional extension as being precisely restricted to the world-line (not tube)  $\gamma$ .

The earliest (implicit) use of distributional energy-momentum in the problem of motion in a relativistic context can be read into the derivations of the geodesic principle made by Myron Mathisson.<sup>28</sup> Infeld did not trade in Einstein’s vacuum-cum-singularity method for distributions until decades later in his (Infeld, 1954, 1957, Infeld & Plebanski, 1960).<sup>29</sup> Even still Infeld (1957, p399) characterized his reluctance about such a shift by noting that though it is technically “unfaithful” to “Einstein’s idea of not using the energy momentum tensor,” the introduction of a distributional energy-momentum tensor can be excused by the fact that it “tremendously simplifies the entire deduction of the equations of motion.”

The key to distribution proofs involves establishing a variational principle for integrals

---

<sup>27</sup>Since smooth tensor fields  $\mathcal{T}_2^0(\mathcal{M})$  are locally integrable and so have a natural embedding in the space of tensor distributions  $\mathcal{D}'_2(\mathcal{M})$ , using variational techniques, this relation trivially entails equations (3.1.1) in cases where the respective fields are smooth (see appendix A).

<sup>28</sup>As we will discuss in section 3.3, Mathisson’s technique involved deriving a variational principle for the integral of an expansion (in gravitational multipole moments) of the energy-momentum tensor field, from which motion can then be deduced. In particular, in (1937, 1940) Mathisson casually shows that applying the principle only to the lowest order term in the expansion (because such a tensor field might be representative of a spinless “point particle”) entails the geodesic equation. Though Mathisson does not make explicit use of distributions in these 0<sup>th</sup>-order derivations, such an appeal can naturally be read into this technique as was done later by Havas & Goldberg (1962). See also (Tulczyjew, 1959) for a distributional reconstruction of Mathisson’s work.

<sup>29</sup>Infeld and Mathisson were colleagues in Poland when Mathisson had been developing the work from which his derivations follow. Infeld, who was familiar with the relevant papers, later conceded that “at the time” he had not understood Mathisson’s (significantly more advanced) methods (Infeld, 1968, p204).

### 3.2. SINGULARITY PROOFS

of the energy-momentum tensor for which, in the special case of energy-momentum tensor *distributions* concentrated on a world-line, the geodicity of the path is entailed. Specifically, assume one is able to establish that for all smooth co-vector fields  $\xi_b$  with compact support:

$$\int_{\mathcal{M}} T^{ab} \overset{g}{\nabla}_a \xi_b \text{vol}_g = 0 \quad (3.2.2)$$

where  $\text{vol}_g$  is the volume element for some metric  $g_{ab}$  and  $\overset{g}{\nabla}_a$  is the derivative operator compatible with  $g_{ab}$ . It follows from (3.2.2) that if  $T^{ab}$  has (distributional) support restricted to a timelike curve  $\gamma$  in some neighborhood around it, then  $\gamma$  is a geodesic of  $g_{ab}$ .<sup>30</sup> So letting  $T^{ab}$  represent a point particle whose world-line is given by  $\gamma$  in that region of the manifold, we might interpret the result as saying that “point particles can only have a geodesic world-lines.”

Since condition (3.2.2) is sufficient for such a distributional representation of a point particle to follow only geodesic world-lines, it is worth considering how such a variational principle can be established in general relativity. Heuristically, we might first note that for *smooth* tensors, it is (in a sense) a *purely mathematical* consequence of the Bianchi identities that the total divergence of the Einstein tensor defined by (3.1.2) vanish (i.e. that  $\nabla_a G^{ab} = \mathbf{0}$ ). Hence, Einstein’s original equations (3.1.1) immediately give us that

$$\overset{g}{\nabla}_a T^{ab} = \mathbf{0} \quad (3.2.3)$$

referred to as *the conservation condition*, which holds for any smooth solution  $(\mathcal{M}, g_{ab}, T^{ab})$  to (3.1.1). Condition (3.2.2) follows for such smooth solutions from (3.2.3) by simply contracting

---

<sup>30</sup>The full proposition is given with a proof in appendix B .

### 3.2. SINGULARITY PROOFS

with arbitrary test co-vector fields  $\xi_b$  and then integrating over the entire manifold.<sup>31</sup>

The problem is that we want (3.2.2) to hold not just for smooth solutions to (3.1.1) but for *distributional* solutions to (3.2.1) as well. And in fact, the Bianchi identities do not automatically hold for distributional Einstein tensors for every solution of the generalized field equations (3.2.1).<sup>32</sup> Hence, *the conservation equation (3.2.3) does not automatically follow in the generalized case of tensor distributions*. But both Mathisson’s implicit distribution result and Infeld’s explicit one overlook this nuance, conflating the distinction between solutions to (3.1.1) and (3.2.1) and then inferring that (3.2.3) (and so (3.2.2)) automatically follows even for distributional sources.<sup>33</sup>

In 1974, Jean-Marie Souriau developed his own “proof” of the geodesic principle by making use of distribution techniques, which again (essentially) take advantage of condition (3.2.2). Unlike earlier attempts, however, Souriau justifies the condition not through the Bianchi identities, but by (rather ingeniously) formulating a “variational” method of expressing the condition that the field equations must be generally covariant (now referred to in the literature as *Souriau’s (local) covariance condition*)(Souriau, 1974).<sup>34</sup> In the case of the generalized field equations (3.2.1), such *Souriau covariance* easily reduces to condition (3.2.2).

Though Souriau’s method is able to avoid the particular invalidity of his predecessors’

---

<sup>31</sup>In fact, as discussed in appendix A, if the connection is smooth, the (covariant) derivative of a distributional  $T^{ab}$  is calculated precisely by negating the left side of equation (3.2.2). Hence, condition (3.2.2) is the natural generalization of the classical conservation condition (3.2.3) and for this reason is sometimes referred to as the *generalized conservation condition*.

<sup>32</sup>Most importantly, they do not automatically hold for the important class of GT-regular solutions discussed in appendix C (see (Geroch & Traschen, 1987, p1020)).

<sup>33</sup>See (Infeld, 1957, §4) and (Mathisson, 1937, §2). See also the explicitly distributional reconstruction of Mathisson’s demonstration in (Havas & Goldberg, 1962, §2). This equivocation can still occasionally be found in introductory texts offering what might be interpreted as “heuristic” derivations of the principle (see e.g. (Hobson *et al.*, 2006, pp188-9)).

<sup>34</sup>The application of Souriau’s local covariance condition was further developed in both relativistic and non-relativistic contexts in (Guillemin & Sternberg, 1978, 1990, Sternberg, 1978, 1985*b,a*, 1999) and (Duval & Künzle, 1978, 1984) respectively.

### 3.2. SINGULARITY PROOFS

arguments, his result is faced with an even more general threat to the use of distribution techniques in Einstein's theory. In order to understand what goes wrong, observe that condition (3.2.2) does not just express a restriction on  $T^{ab}$ , but rather it expresses a restriction on the energy-momentum tensor field (or distribution) in relation to a metric  $g_{ab}$ . Hence, condition (3.2.2) (and Souriau's subtly more general covariance condition) define a subset of *ordered pairs* of symmetric tensor fields or tensor distributions  $(g_{ab}, T^{ab})$  defined on a manifold  $\mathcal{M}$ .<sup>35</sup> Let us refer to such pairs as *Souriau pairs*. Of course, solutions to the generalization of Einstein's field equations (3.2.1) can also come in ordered pairs defined on  $\mathcal{M}$ . And Souriau's covariance principle is supposed to establish that *for any solution  $(g_{ab}, T^{ab})$  to (any) generally covariant field equations (such as (3.2.1)),  $(g_{ab}, T^{ab})$  will constitute a Souriau pair*. So the logic works as follows:

1. By Souriau's covariance argument, if  $(g_{ab}, T^{ab})$  is a solution to (any) generally covariant field equations (such as (3.2.1)), then the pair satisfies condition (3.2.2).
2. And by proposition B.2,<sup>36</sup> if  $(g_{ab}, T^{ab})$  satisfies condition (3.2.2), and  $T^{ab}$  is concentrated as a distribution onto a timelike world-line, then it must be a geodesic of  $g_{ab}$ .

- 
3. Hence, if  $(g_{ab}, T^{ab})$  is a solution to (3.2.1), and  $T^{ab}$  is concentrated as a distribution onto a timelike world-line, then it must be a geodesic of  $g_{ab}$ .

Though this argument is valid, the antecedent of line 3 renders the conclusion (essentially) vacuous. The reason for this was ultimately demonstrated by Geroch & Traschen (1987,

---

<sup>35</sup>Moreover, in the context of linear distributions,  $g_{ab}$  must be non-degenerate, and (at minimum) it must have a smooth connection wherever  $T^{ab}$  behaves singularly (as a distribution).

<sup>36</sup>See appendix B.



### 3.2. SINGULARITY PROOFS

Thm. 1) in which they show that the only reasonable<sup>37</sup> solutions  $g_{ab}$  to the equations (3.2.2) cannot have sources  $(T^{ab})$  with support concentrated on a one-dimensional curve.<sup>38</sup> Hence, there is no (reasonable) distributional solution  $(g_{ab}, T^{ab})$  to Einstein’s generalized field equations (3.2.1) such that  $T^{ab}$  can also be concentrated onto a timelike world-line.

Geroch and Traschen’s theorem hence serves as an effective death nail for attempts to deduce the geodesic principle from Einstein’s field equations by using singular models of one-dimensional “point” particles. As a result we seem to be left with the following option: We could conduct deductions from *inexact* (namely linearized) field equations.<sup>39</sup> It is possible to deduce exact geodesic motion in this case, *but not from Einstein’s actual field equations*. Alternatively, we might move to so-called “second way” proofs attempting to deduce geodesic motion from Einstein’s (non-vacuum) field equations (3.1.1) using smooth energy-momentum tensors with four-dimensional support to represent our geodesic following objects.<sup>40</sup> In the next two sections we will review the major strategies that have been used in such “second way” deductions. As we shall see, by moving to a context of extended models of massive bodies, much more freedom in the behavior of the object is introduced, ultimately leading to the deduction of *non-geodesic* motion in generic cases. It will turn out that these additional modes of freedom can be later reduced by appealing to certain limiting cases, but this has the detrimental result of either bringing us back to the context of singularity proofs, or to

---

<sup>37</sup>Recall, according to (3.2.1) the energy-momentum tensor is equated with the Einstein curvature tensor, and the curvature tensor in turn depends on the metric and its derivatives. Hence, the metric and its derivatives must meet certain minimal conditions on their integrability in order for “integration” (i.e. the action) of the Einstein tensor (and so the energy-momentum tensor) to make sense as a well defined tensor distribution. Geroch and Traschen define a class of metrics now called *GT-regular metrics* designed to meet such conditions so that the energy-momentum tensor distributions determined by these metrics can make sense. See appendix C for the precise definition of this class of metrics and a brief discussion of why this class in particular constitutes the appropriate class of “reasonable solutions” to (3.2.2).

<sup>38</sup>They prove that distributional sources must have support of co-dimension no greater than 1 in  $\mathcal{M}$ .

<sup>39</sup>The Geroch-Traschen proof crucially depends on the non-linear dependence of Einstein curvature on the metric. See appendix C for discussion.

<sup>40</sup>A third possibility involving neither the original field equations (3.1.1) nor the (linear) distributional generalization (3.2.1) that might be developed is considered in Appendix C (see note 4).

### 3.3. $0^{TH}$ -ORDER PROOFS

limiting representations in which (contrary to the canonical account) the “gravitating” body simply vanishes.

## 3.3 $0^{th}$ -Order Proofs

Einstein’s field equations have an initial value formulation. Under suitable conditions, this problem can be well posed so that we might use the field equations to deduce how a given tensor field defined on a particular hypersurface evolves over time (viz the domain of dependence of that hypersurface). Though this works in principle, such a program is far more easily said than done in most cases and often numerically rather than analytically. As a consequence if physicists wished to predict, say, the motion of celestial bodies, it was more practical to figure out a way to approximate the structure of the bodies and the metric field affected by their presence by expanding both of these fields through various procedures and then dropping some of the “higher order” terms.<sup>41</sup> After the respective fields have been suitably simplified in this way, the physicist can take steps to determine the expected “approximate” paths of such bodies. For the most part *the resulting paths are not geodesic*. However, when *all* of the higher order terms of the tensor field representing the energy-momentum of a body (and the tensor field representing the body’s effect on the metric) are dropped, it is the case that one is able to “deduce” geodesicity from the reduced equations. Since these proofs share the feature that all higher order terms accounting for the energy-momentum of the body

---

<sup>41</sup>Aside from historical computational hurdles, there is the epistemological motivation investigated in chapter 2 for why such approximation techniques might be advantageous. Models of gravitating bodies whose higher order effects are ignored may be used to develop an  $\epsilon$ -faithful representation through which we can draw conclusions about bodies with higher order terms that are non-vanishing but ( $\epsilon$ )-small in some relevant way. In such cases, the former (higher-order term free) models may serve anchor models for the relevant  $\epsilon$  to  $\delta$  deductions establishing the  $\epsilon$ -fidelity of these representations. As we shall elaborate in chapter 4 when we discuss the geodesic universality thesis, geodesic models can be used to help draw such conclusions about the bulk behavior of general massive bodies that may be obfuscated by attending to every detail as in the initial value formulation.

### 3.3. $0^{TH}$ -ORDER PROOFS

(and its effects on the metric) must be dropped in order to ensure geodicity, this class of deductions will be referred to as  $0^{th}$ -order proofs. The overwhelming majority of geodesic “derivations” in the literature can be classified as belonging to this family of proofs. The explanation for this pattern is entirely pragmatic: expansion methods are an efficient way of generating approximations of the motion of bodies under the influence of relativistic effects within particular margins of error.<sup>42</sup> So,  $0^{th}$ -order proofs are quite abundant in the literature, but ironically only thanks to the desire for approximations of motion accurate to degrees *higher* than  $0^{th}$ -order, the lowest order geodesic deductions being offered as a kind of afterthought or perfunctory check.<sup>43</sup>

By far, the earliest concerted attempts to approximate motion in relativity theory by expanding the energy-momentum of a body can be found in the works of Mathisson (1937, 1940). To understand the sense in which Mathisson “expanded” the energy-momentum tensor, consider a (timelike) world-tube  $\mathcal{W}$  in a relativistic spacetime  $(\mathcal{M}, g_{ab})$ .<sup>44</sup> We then consider a symmetric (locally integrable) tensor field  $T^{ab}$  with support contained in  $\mathcal{W}$ . This tensor field can be interpreted as representing the energy-momentum of a body moving along in the world-tube  $\mathcal{W}$ . At this point, Mathisson takes advantage of the fact that we can understand the properties of such a field by considering the following linear operation

---

<sup>42</sup>Bursts of progress in such techniques often appear to be motivated by concomitant instrumental advances demanding further accuracy. For example, the advances in approximating “self-force” effects in the last dozen years seem to have been originally motivated by the promise of gravitational wave detectors (see (Quinn & Wald, 1999)).

<sup>43</sup>Because of such abundance, in this section we will not attempt a comprehensive review of all major attempts. Instead, we focus only on a few examples paradigmatic of the general methods of expansion techniques recovering geodesic motion (at the lowest order).

<sup>44</sup>Since our purpose in this section is primarily illustrative, in the following discussion it is assumed that the spacetime is simply connected, orientable and that the metric is smooth. Some of these constraints might be relaxed, but as discussed above in section 3.2.2, doing so can lead to serious complications. Though Mathisson’s work was quite sophisticated for his time, there are a number of mathematical ambiguities in his original formulation that we will not dwell on here, especially since such infelicities were eventually rectified by Dixon (see below). For example, the sense in which  $\mathcal{W}$  is timelike is made precise by Dixon through a construction that involves joining a particular set of local, convex, disjoint, hypersurfaces with compact closure that are normal to a timelike “baseline” curve.

### 3.3. $0^{TH}$ -ORDER PROOFS

defined on arbitrary smooth tensor fields  $\phi_{ab}$  also compactly supported in  $\mathcal{W}$ :

$$\langle T^{ab}, \phi_{ab} \rangle \mapsto \int_{\mathcal{M}} T^{ab} \phi_{ab} \text{vol}_g \quad (3.3.1)$$

Picking an arbitrary smooth timelike curve  $I \ni s \mapsto \gamma(s) \in \mathcal{W}$  parametrized by proper time  $s$ , and letting  $\Sigma(s)$  be a local foliation of  $\mathcal{W}$  parametrized by  $s$  and meeting certain orthogonality conditions with respect to  $\gamma$ , Mathisson proceeds to show that by integrating across the  $\Sigma(s)$ 's the action (3.3.1) can be equivalently approximated by a series of integrals along  $\gamma$  as follows:

$$\int_{\mathcal{M}} T^{ab} \phi_{ab} \text{vol}_g = \int_{\gamma[I]} \left( I_0^{ab} + I_1^{abm_1} \nabla_{m_1} + I_2^{abm_1m_2} \nabla_{m_2} \nabla_{m_1} + \dots \right) \phi_{ab} ds \quad \forall \phi_{ab} \quad (3.3.2)$$

Where the tensor fields  $I_n^{abm_1 \dots m_n}$  satisfy certain symmetry conditions and orthogonality conditions with respect to  $\gamma$ . These tensor fields represent the  $2^n$ -gravitational-multipoles of the body  $T^{ab}$ . The final move to arrive at Mathisson's variational equation of motion is to let  $\phi_{ab} = \nabla_b \xi_a$  for arbitrary smooth, compactly supported co-vector fields  $\xi_a$ . But in these cases, the left side of (3.3.2) takes the form of the left side of the generalized conservation condition (3.2.2), which as we discussed above is equivalent to the traditional conservation condition for a smooth  $T^{ab}$  and metric. Based on this reason, Mathisson sets the left side of (3.3.2) to 0, giving us the final form of his *variational equation of motion*:

$$0 = \int_{\gamma[I]} \left( I_0^{ab} + I_1^{abm_1} \nabla_{m_1} + I_2^{abm_1m_2} \nabla_{m_2} \nabla_{m_1} + \dots \right) \nabla_b \xi_a ds \quad \forall \xi_a \quad (3.3.3)$$

Mathisson was able to use this variational equation, expanded to the first (dipole) and to

### 3.3. $0^{TH}$ -ORDER PROOFS

the second (quadrupole) terms to generate explicit approximate equations of motion for a “small test body” with angular momentum and spherical asymmetries. These equations were (much later) derived in the better known (Papapetrou, 1951), using hyperbolic coordinates and somewhat similar expansion techniques, and are sometimes referred to as the *Mathisson-Papapetrou equations*. In the late 60’s W. G. Dixon took up the Mathisson project, eventually producing a fantastic series of papers (Dixon, 1964, 1967, 1970b,a, 1973, 1974, 1975), which rigorously put what he called the *multipole approximation* technique on fully matured mathematical footing.

The geodesic result (already mentioned in section 3.2) comes when we drop all higher order terms in the expansion of equation (3.3.3) to get the variational constraint:

$$0 = \int_{\gamma[I]} I_0^{ab} \nabla_b \xi_a ds \quad \forall \xi_a \quad (3.3.4)$$

From this constraint one can then deduce that if  $I_0^{ab} \neq \mathbf{0}$  on  $\gamma[I]$  then the curve is a geodesic of the metric compatible with the connection  $\nabla_a$ .<sup>45</sup> So since (*ex hypothesi*)  $I_0^{ab}$  is supposed to be the only (significant) contribution to the energy-momentum of the body and our initial choice of  $\gamma$  was arbitrary, we can think of this result as telling us that certain “suitable”<sup>46</sup> timelike  $\gamma$ ’s contained in such a body must be geodetic. The victory is rather Pyrrhic, however, because approximations involving even one non-vanishing multipole term will no longer describe geodesic motion, instead predicting a kind of “wobbling” behavior inside  $\mathcal{W}$ .

---

<sup>45</sup>A deduction (somewhat different from that of Mathisson’s original proof) is contained in the proof of proposition B.2 in appendix B (condition (3.3.4) is equivalent to condition (B.2), which in the course of the proof is shown to entail geodeticity, when we also assume that the body is non-vanishing on  $\gamma$ ).

<sup>46</sup>The *suitable*  $\gamma$  must be a curve with respect to which the higher order terms  ${}_n I^{abm_1 \dots m_n}$  can be dropped. Intuitively this can be thought of as a timelike axis of symmetry with respect to all of the multipole moments of the body. As we shall see, when we expect perfect geodeticity (i.e. that all the higher order terms can be dropped because with respect to such a  $\gamma$  axis they actually vanish) this “suitability” condition becomes prohibitively strong.

### 3.3. $0^{TH}$ -ORDER PROOFS

In order to dodge this Pyrrhic foil, we might search for ways to justify the use of constraint (3.3.4) as an exact (or at least stable) description satisfying the canonical interpretation. One possibility is to justify the constraint by insisting that, if the body is a point particle with only timelike extent, then the particle’s lack of spatial extent means that all higher order multipole terms must vanish, making the only curve left in the support of the point particle relevantly “suitable” and hence geodetic. This approach validates the inference to constraint (3.3.4) from Mathisson’s full variational equation (3.3.3), but it has the unfortunate side effect of turning the deduction into a variety of singularity proof which as we saw in section 3.2 is incoherent with Einstein’s field equations.

A second strategy is to accept that the body has spacelike extent, but then suggest that (3.3.4) holds of a certain “conceivable” type of material body, namely, one that is perfectly symmetrical about some timelike  $\gamma$  with respect to *every* higher order multipole moment. Though by construction such a representation would reduce the equation (3.3.3) to constraint (3.3.4), the proof is far from general. Such a “conceivable” body is not possible according to just about any serious theory of matter considered by physicists. For example, such extreme symmetry constraints would require that the body could not be composed of atomic or molecular constituents for such inhomogeneities would necessitate spherical asymmetries in the distribution of the body’s energy-momentum.<sup>47</sup> Moreover, such restrictions necessary to generate geodesic motion are highly unstable under perturbations. Any change in the angular momentum (or any other multipole moment) would break the symmetry needed to exactly recover constraint (3.3.4). Though we might wish to identify these perfectly symmetrical constructs as a kind of “idealization” in Einstein’s theory, *the idealization is degenerate*

---

<sup>47</sup>The body would need to consist of something like the metaphysically curious homogeneous material famously considered by Saul Kripke. In the context of Kripke’s spinning disk, it is ironic that one of the constraints that we are explicitly placing on such a material body is that it would have to be completely “spin free” in the specific sense that  ${}_1I^{abm_1} = \mathbf{0}$ .

### 3.3. $0^{TH}$ -ORDER PROOFS

with respect to recovering precise geodesic motion: the path becomes non-geodesic under *arbitrarily small* perturbations of the energy-momentum.<sup>48</sup>

Before moving on to the final class of geodesic deductions, it is worth pointing out that by invoking a multipole approximation of the energy-momentum, the body considered is no longer coupled to the metric in accordance with Einstein’s equations (3.1.1). In other words, multipole approximations (even ones that do involve some higher order terms such as the Mathisson-Papapetrou equations) are guilty of adopting test body approximations that have not been explicitly justified. They ignore the so-called “back reaction” or “self-force” effects that the body has on the metric.<sup>49</sup> Hence, when considering the path of extended bodies (i.e. “second way” deductions) we must not only worry about correcting for the possibility of unjustified test body approximations facing earlier proofs, but now must also manage the “spin effect” corrections to geodesic motion resulting from the internal degrees of freedom available to an energy-momentum tensor with spacelike extent. In the next section we will consider the final family of deductions, which attempts to manage such deviations from geodicity by conducting certain limit operations. As we shall see, it is only by taking limits that appropriately manage both of these sources of non-geodicity that success is achieved. Unfortunately for the canonical view, this will also mean that the body must vanish completely before we can recover such geodicity.

---

<sup>48</sup>The degenerate instability that results when we try to recover *exact geodesic motion* in accordance with the canonical interpretation is analogous to the examples discussed in chapter 2, where accounts of mathematical representation aimed at narrowly matching very specific constraints were shown to lead to Plato’s problem. As indicated in the opening of this section, we shall argue in chapter 4 that it is possible to eliminate (or at least avoid) this kind of (pathological) degeneracy, once we understand the role of such geodesic “idealizations” as anchor models used to establish an  $\epsilon$ -faithful account of the near geodesic behavior of suitable massive bodies.

<sup>49</sup>Significant advances have been made in (Mino *et al.*, 1997, Quinn & Wald, 1997, Gralla & Wald, 2008, Pound, 2010), which approximate the consequences of “self-force” effects as first order perturbations in the background metric. Though these methods are not without their own difficulties (particularly when it comes to justifying what is referred to as the “Lorentz gauge relaxation”) because these self-force effects lead to violations of geodesic motion, these complications will not be elaborated here. For a nice introduction to these issues see (Wald, 2011).

## 3.4 Limit Operation Proofs

The strategy behind the final family of *limit operation proofs* is to avoid the complications arising from investigating the motion of “true” point particles with extent restricted precisely to one-dimensional timelike curves by instead considering sequences of energy-momentum representations of particles whose spacelike extent is confined to increasingly smaller neighborhoods of those curves. We can think of these infinite sequences of tensor fields as representing particles with *arbitrarily small (but non-zero) spacelike extent* in the sense that no matter how “narrow” we might want the “particle” to be, eventually the sequence will list tensor fields with support entirely confined in such a narrow region. The strategy then is to show that if such a sequence of fields can be constructed for a given curve  $\gamma$ , then  $\gamma$  must be a (timelike) geodesic, allowing one (roughly) to claim that “arbitrarily small particles must follow (timelike) geodesics.”

### 3.4.1 Geroch-Jang Particles

In contrast to the overabundance of  $0^{th}$ -order derivations, the class of limit proofs consists primarily of two elegant results.<sup>50</sup> The first result by [Geroch & Jang \(1975\)](#) considers sequences of ostensible energy-momentum tensor fields of ever narrowing support. More

---

<sup>50</sup>The self-similarity limit operations done by [Gralla & Wald \(2008\)](#) can (in part) also be classified as an enhancement of these limit proof strategies, though they then proceed to employ some of the expansion techniques discussed in section 3.3. Hence, ([Gralla & Wald, 2008](#)) appears to constitute a kind of borderline case between the two families. The work done far earlier by [Robertson \(1937\)](#) might also be considered a kind of proto-limit operation proof attempt, in that he follows the general strategy of considering the limiting behavior of an extended “corpuscle” as the spatial extent goes to 0. In contrast, the limits taken by [Infeld & Schild \(1949\)](#) in the “IS-particle” constructions discussed in section 3.2.1 would determinately not count as a member of the family of limit operation proofs we are considering in that (aside from the ill-defined convergence issues already discussed) for each of the sequence entries the test body is represented by a singularity rather than a smooth tensor field representation of an *extended* object. This distinction between considering the limiting behavior of sequences of *extended* bodies and merely attempting to appeal to limits in the course of the demonstration is significant, and the former more restrictive characteristic is required for our present classification (cf. [Havas \(1989, p254\)](#) who seems to overlook the significance this distinction).



### 3.4. LIMIT OPERATION PROOFS

precisely, Geroch and Jang's theorem can be formulated as follows:

**Theorem 3.1. (Geroch-Jang, 1975)** *Let  $\gamma : I \rightarrow \mathcal{M}$  be a smooth curve in Lorentzian spacetime  $(\mathcal{M}, g_{ab})$ . Suppose that given any open neighborhood  $\mathcal{O}$  of  $\gamma[I]$ , there exists a smooth symmetric tensor field  $T_{ab}$  defined on  $\mathcal{M}$  such that for all points  $p \in \mathcal{M}$ :*

1.  $T_{ab}$  has non-vanishing support contained in  $\mathcal{O}$ ,
2. For all timelike  $\xi^a$ :  $T_{ab}\xi^a\xi^b \geq 0$  and if  $T_{ab} \neq \mathbf{0}$ , then  $(T_{ac}\xi^a)(T^c_b\xi^b) > 0$ ,
3.  $\overset{g}{\nabla}_a T^a_b = \mathbf{0}$ ,

then  $\gamma[I]$  is the image of a timelike  $g_{ab}$ -geodesic.

In order to illustrate the significance of these conditions, consider any set of nested neighborhoods  $(\mathcal{O}_i)_{i \in \mathbb{N}}$  that becomes arbitrarily narrow around the curve  $\gamma$  as  $i \rightarrow \infty$ . Next consider a sequence of tensor fields  $(T_{ab})_{i \in \mathbb{N}}$  such that each  $T_{ab}$  satisfies conditions 2 and 3, and for each  $i$ ,  $T_{ab}$  satisfies condition 1 for the neighborhood  $\mathcal{O}_i$ . Let us refer to such a sequence of tensor fields  $(T_{ab})_{i \in \mathbb{N}}$  as a *Geroch-Jang* or *GJ-particle*. The existence of a GJ-particle sequence for an arbitrary sequence of nested neighborhoods tightening around  $\gamma$  is equivalent to the satisfaction of the conditions of theorem 3.1.

Let us consider what each of the conditions says about GJ-particles. First, condition 1 establishes the arbitrary smallness characteristic of GJ-particles. Since the nested neighborhoods  $(\mathcal{O}_i)_{i \in \mathbb{N}}$  become arbitrarily narrow, no matter how tight around  $\gamma$  we demand that the world-tube of the particle be confined,  $T_{ab}$  will eventually (for sufficiently large  $i$ ) stay that close (or closer). If we interpret the symmetric field  $T_{ab}$  as an energy-momentum tensor, then condition 2 says that from the perspective of any observer the matter-energy in that part of the universe (a) is non-negative and (b) only flows in timelike directions (it doesn't go as fast

### 3.4. LIMIT OPERATION PROOFS

or faster than the speed of light). So, roughly speaking, the theorem is only telling us about the behavior of a certain kind of body that is of positive mass (as opposed, for example, to photons, which we don't expect to follow timelike geodesics anyway or the kind of body we might wish to associate, say, with non-classical negative energy solutions). Hence, such conditions restricting the kind of matter-energy that can constitute a GJ-particle appear appropriate for the sort of material body about which we expect the principle to be relevant. However, it is worth observing that condition 2 *does not follow from Einstein's field equations*, and hence constitutes an additional assumption that must be obeyed in order to get Geroch and Jang's geodesic result.<sup>51</sup>

Condition 3 is the familiar conservation condition (3.2.3), which (as discussed in section 3.2.2) follows directly from Einstein's original field equations for all smooth solutions. Condition 3 is the primary reason why one might say that theorem 3.1 constitutes a “deduction” of the geodesic principle from Einstein's field equations. *If  $(g_{ab}, T_{ab})$  were a solution to Einstein's field equations, then condition 3 would be automatically satisfied for  $T_{ab}$ .* This might allow us (roughly) to characterize Geroch and Jang's result by claiming that “arbitrarily small bodies of positive mass that obey Einstein's field equations must follow geodesics.”<sup>52</sup>

---

<sup>51</sup>A recent discussion of this logical independence can be found in (Malament, 2012), along with a demonstration that the existence of an “almost” GJ-particle (viz ones that satisfy the first and third but not the second condition) fails to ensure that  $\gamma$  is a geodesic. See also (Weatherall, 2011) in which it is shown that condition 2 (as opposed to a slightly weaker energy condition originally used by Geroch and Jang) is necessary.

<sup>52</sup>A nuance worth noting that is imposed by condition 3 involves the question of electromagnetic (or other non-gravitational) field effects. One might think that this phrasing of the principle is too strong: though *neutral* massive bodies are supposed to follow geodesics, charged bodies under the influence of an electromagnetic field should not. Of course, if there is an electromagnetic field to influence our GJ-particle it would supply further energy-momentum ( $T_{ab}^{(EM)}$ ) in the neighborhoods of  $\gamma$  (indeed, electromagnetic energy-momentum that should not stay confined to arbitrary neighborhoods of  $\gamma$ ). So if, for example, we were talking about a charged body  ${}_i T_{ab}$  interacting with an electromagnetic field  $T_{ab}^{(EM)}$  near  $\gamma$ , then Einstein's field equations only ensure that the *total* energy-momentum is conserved (e.g.  ${}^g \nabla_c g^{ca} ({}_i T_{ab} + T_{ab}^{(EM)}) = 0$ ). This means that when we interpret the GJ-particle entries  ${}_i T_{ab}$  as representing the matter-energy flow of (small, massive) bodies, condition 3 can be thought of as requiring that the bodies are “free” in the sense that their energy-momentum does not change due to interactions with other non-gravitational energy carrying fields in the neighborhood of  $\gamma$ . Hence, we might paraphrase the result even more appropriately by saying

### 3.4. LIMIT OPERATION PROOFS

Unfortunately, this claim reveals a significant complication. Assume that the  $T_{ab}$  represent the entire contribution to the source side of Einstein's field equations (3.1.1) in their respective neighborhoods of  $\gamma$ . In this case, they each should be having a non-zero perturbative effect ( $\epsilon_{ab}$ ) on the metric of the spacetime manifold. The problem is that since the support of the GJ-particle is non-vanishing but continually shrinking down in spacelike extent, this non-zero perturbative effect  $\epsilon_{ab}$  will not be stable in every region of spacetime even for sufficiently large  $i$ . That is to say, if solutions to the respective  $T_{ab}$  of a GJ-particles take the form  $g_{ab} = g_{ab} + \epsilon_{ab}$ , the perturbation field  $\epsilon_{ab}$  must vary for some larger  $i$  value. This means in particular that  $\epsilon_{ab}$  cannot vanish for arbitrarily large  $i$ , so no matter how far out in the sequence we look the remaining  $T_{ab}$ 's will never all couple to the metric  $g_{ab}$  in accordance with Einstein's field equations. Hence, the constraints placed on the existence of a GJ-particle do not prevent differences between (a) the geodesic structure of a spacetime in which a GJ-particle entry obeys Einstein's field equations and (b) the geodesic structure of the background metric  $g_{ab}$  according to which  $\gamma$  actually counts as a geodesic. In other words, geodesics of  $g_{ab}$  will not necessarily remain geodesics of the spacetimes with GJ-particles in them. We can think of imposing a GJ-particle in neighborhoods of  $\gamma$  as having a kind of "bending" effect on  $\gamma$ , ruining its geodeticity.

So while it is nice to know that GJ-particles must follow geodesics of *some* spacetime, strictly speaking *the theorem does not ensure that GJ-particles must follow geodesics of the spacetime(s) in which they might actually exist (at least not without violating Einstein's field equations)*. Though Geroch and Jang's theorem is appealing with respect to its mathematical elegance and certain aspects of its representational fertility, in this form it is guilty of relying on a test body approximation that is ultimately left unjustified by the conditions. That being

---

that "arbitrarily small free bodies of positive mass must follow geodesics." (But see next.)

### 3.4. LIMIT OPERATION PROOFS

said, reflection on the theorem leaves one with the sense that “if only the perturbative effect  $\epsilon_{ab}$  could be controlled somehow as the GJ-particle shrinks down in size, then we might at least be able to say that the spacetime metrics  $g_{ab}$  coupling to our GJ-particle entries should ‘come close’ to the background metric  $g_{ab}$ .”<sup>53</sup> In the next section, we will consider a second limit operation proof achieving just this sort of result.

#### 3.4.2 Ehlers-Geroch Particles

In 2004, Ehlers, who had evidently been concerned by the Geroch-Jang “test body approximation” for nearly three decades (see note 8), collaborated with Geroch to develop a second result that accommodates for the “geodesic bending” effects of GJ-like particles. If Geroch and Jang’s theorem approaches things from the “source side” of Einstein’s field equations (3.1.1), then the Ehlers & Geroch (2004) result can be said to approach things from the geometry side of the equations. Specifically, instead of considering sequences of energy-momentum tensors, they consider sequences of *metrics* which converge in an appropriate way to a background metric and whose energy-momentum sources simultaneously satisfy (essentially) the same conditions as those placed on GJ-particles. Their result can be formulated as follows.<sup>54</sup>

**Theorem 3.2. (Ehlers-Geroch 2004)** *Let  $\gamma : I \rightarrow \mathcal{M}$  be a smooth timelike curve in Lorentzian spacetime  $(\mathcal{M}, g_{ab})$ . Suppose that for any sufficiently small closed neighborhood  $\mathcal{K} \subset \mathcal{M}$  of  $\gamma[I]$  there exists a sequence of smooth Lorentzian metrics  $g_{ab}$  defined on  $\mathcal{K}$  such that for all points  $p \in \mathcal{K}$ :*

---

<sup>53</sup>Note, for any GJ-particle sequence  $({}_i T_{ab})_{i \in \mathbb{N}}$ , there exists a second GJ-particle sequence  $({}_i \hat{T}_{ab})_{i \in \mathbb{N}}$  whose matter-energy density vanishes arbitrarily quickly as  $i \rightarrow \infty$ . (Just define  ${}_i \hat{T}_{ab} := (\alpha_i)({}_i T_{ab})$  for each  $i$  where  $(\alpha_i)_{i \in \mathbb{N}}$  is a sequence of scalars converging to 0 with suitable quickness.)

<sup>54</sup>The theorem as stated originally in (Ehlers & Geroch, 2004) is slightly stronger than the following version: their result still goes through if the strict inequality in condition 2 is weakened to allow for equality as well. Of course the theorem as stated is an immediate consequence of the slightly stronger version. The difference is inconsequential to our current discussion.

### 3.4. LIMIT OPERATION PROOFS

1. For all  $j$ :  $G_{ab}$  has non-vanishing support contained in the interior of  $\mathcal{K}$ ,
2. For all  $j$  and timelike  $\xi^a$ :  $G_{ab}\xi^a\xi^b \geq 0$  and if  $G_{ab} \neq \mathbf{0}$ , then  $g_j^{bd}\left(G_{ab}\xi^a\right)\left(G_{cd}\xi^c\right) > 0$ ,
3. The  $g_{ab} \rightarrow g_{ab}$  as metrics in  $\mathcal{C}^1(\mathcal{K})$  as  $j \rightarrow \infty$ ,<sup>55</sup>

where  $G_{ab}$  is the Einstein curvature tensor determined by  $g_{ab}$ , then  $\gamma[I]$  is the image of a  $g_{ab}$ -geodesic.

As with GJ-particles, we might illustrate the content of the theorem by considering an arbitrary set of nested neighborhoods  $(\mathcal{K}_i)_{i \in \mathbb{N}}$  converging down around  $\gamma[I]$ . For these neighborhoods, we can now consider the double indexed sequences of smooth metrics  $(g_{ab})_{i,j \in \mathbb{N}}$  and corresponding curvature tensors  $(G_{ab})_{i,j \in \mathbb{N}}$  defined for each  $i$  on  $\mathcal{K}_i$ . The latter sequence of curvature tensors  $(G_{ab})_{i,j \in \mathbb{N}}$  can then be identified via Einstein's equations as a sequence of energy-momentum tensors  $(T_{ab})_{i,j \in \mathbb{N}}$  which we will call *Ehlers-Geroch* or *EG-particles*. Observe, it follows from conditions 1 and 2 in theorem 3.2, that each  $T_{ab}$  satisfies conditions 1 and 2 of theorem 3.1 in  $(\mathcal{K}_i, g_{ab})$ . Moreover, because each  $T_{ab}$  is equal to a smooth Einstein curvature tensor, they automatically satisfy condition 3 of theorem 3.1 in  $(\mathcal{K}_i, g_{ab})$  as well. Hence, EG-particles are quite similar to GJ-particles, the difference with EG-particles is that the further condition 3 ensures that the perturbative effect of EG-particles on the “background metric”  $(g_{ab})$  can be made *arbitrarily small* (in the relevant senses) for sufficiently large  $j$ . Most importantly, the convergence demanded in condition 3 ensures not only that the  $\gamma$  is a geodesic of the background metric  $g_{ab}$ , but also for sufficiently large

---

<sup>55</sup>The  $\mathcal{C}^1(\mathcal{K})$  topology  $\tau_{\mathcal{C}^1(\mathcal{K})}$  is defined on the space of ordered pairs of symmetric tensor fields and covariant derivative operators defined for the closed region  $\mathcal{K}$ .  $\tau_{\mathcal{C}^1(\mathcal{K})}$  consists of point-wise neighborhoods of the tensor fields and connections respectively, varying continuously with  $p$  but otherwise arbitrarily in the respective spaces (this can be done explicitly, for example, with the selection of arbitrary positive definite metrics defined on  $\mathcal{K}$ ). We say that  ${}_j g_{ab} \rightarrow g_{ab}$  as metrics in  $\mathcal{C}^1(\mathcal{K})$  if for every neighborhood  $\mathcal{N} \in \tau_{\mathcal{C}^1(\mathcal{K})}$  of  $(g_{ab}, \nabla_a)$ , for sufficiently large  $j$  we have  $({}_j g_{ab}, \nabla_a^{(j)}) \in \mathcal{N}$  where  $\nabla_a$  and  $\nabla_a^{(j)}$  are the unique derivative operators compatible with their respective metrics. Because these operators are uniquely determined by their metrics, explicit reference to them can be suppressed in the articulation of the theorem.

### 3.4. LIMIT OPERATION PROOFS

$j$ ,  $\gamma$  will come arbitrarily close to being a timelike geodesic with respect to the “perturbed” metrics  $g_{ab}^{(i,j)}$ . So not only can EG-particles (like their GJ-particle cousins) be made arbitrarily small in (spacelike) extent around  $\gamma$  for sufficiently large  $i$ , unlike GJ-particles by picking sufficiently large  $j$  we can also control the “geodesic bending” effects resulting from their presence, ensuring that the EG-particle will be shrunken down around a curve that will come arbitrarily close to actually being a geodesic. Hence, Ehlers and Geroch’s result can be characterized by the claim that “arbitrarily small bodies of positive mass come arbitrarily close to following geodesics.”

Unfortunately for the canonical view, the theorem cannot ensure the actual geodicity of  $\gamma$  in the presence of any massive body obeying the field equations. Though by “turning up” the  $i$  and  $j$  indices so to speak, we can make the EG-particle both as narrow and “close to straight” as we want, *we can never ensure actual geodicity for any finite  $j$* . Actual geodicity is only achieved *at the limit* in the spacetime with the metric  $g_{ab}$ . Ehlers and Geroch’s theorem ensures near geodicity in the *approach to the limit*, but if in accordance with the canonical view we are looking to ensure massive bodies following actual geodesics, theorem 3.2 does not do the trick.<sup>56</sup>

An ardent defender of the canonical view might attempt to get around the fact that geodicity is not acquired for  $j < \infty$  by focusing on “the limit case” directly. One problem with this is that though condition 3 establishes convergence of the metrics and an approach to the geodicity of  $\gamma$ , it is insufficient for the convergence of EG-particles to a “limit” energy-

---

<sup>56</sup>Note the similarity to our discussion of the rationallium and phase transition examples considered in section 2.3 above. In all three cases, we have a situation where there exists some relevant more “finely grained” topology with respect to which the series of models comes “close” to having an (qualitatively distinguishable) anchor property (viz having a volume ratio of exactly 2, behaving non-analytically, and following an absolutely geodesic path, respectively). As with the former two examples, our below resolution of these geodesic pathologies will involve relaxing the requirement of precisely meeting said properties, so that we can also take advantage of the closeness afforded by the more finely grained topology.

### 3.4. LIMIT OPERATION PROOFS

momentum field.<sup>57</sup> In general, there is no energy-momentum limit of an EG-particle. In fact, the only way to get convergence of EG-particles in a way that does not allow for new energy-momentum to suddenly appear at the limit (but not before) and likewise does not violate Einstein's field equations at the limit is by having the EG-particle converge to a tensor field that vanishes around  $\gamma$ .<sup>58</sup> In other words, EG-particles don't generally converge, but even the ones that might converge (at least in any acceptable way) either violate Einstein's field equations or vanish. Again the canonical view is left unable to establish actual geodesic motion for massive bodies. Though EG-particles can be said to "come close" to exhibiting geodesic motion, the only way to establish actual geodicity is by violating the field equations or having energy-momentum of the particle completely disappear. As in the case of  $0^{th}$ -order

---

<sup>57</sup>Ehlers and Geroch consider an explicit counterexample sequence of spacetimes whose metrics converge according to condition 3, but whose associated curvature tensors (and so energy-momentum tensors) become divergent. The reason for this possibility is that curvature tensors involve not just derivatives of the metric but also of the connection (see equations (C.1) and (C.2) below), but convergence in the  $\mathcal{C}^1$  topology only ensures closeness of the metric and the connection, but not higher derivatives. They note that ruling out such examples would involve strengthening condition 3 to require convergence in a more restrictive " $\mathcal{C}^2(\mathcal{K})$  topology."

<sup>58</sup>To see why this is the case, let  $\mathcal{K}$  be an arbitrary sufficiently small closed neighborhood of  $\gamma$  from theorem 3.2. If we want to preserve coherence with Einstein's field equations, then any "converging" sequence of energy-momentum tensors  $({}_jT_{ab})_{j \in \mathbb{N}}$  coupled to the metrics  ${}_jg_{ab}$  defined for  $\mathcal{K}$  as in theorem 3.2 will have to converge to some energy-momentum tensor  $T_{ab}$  equal to the Einstein curvature tensor  $G_{ab}$  determined by the background metric  $g_{ab}$ . (Otherwise, it would be the case that the energy-momentum "in the limit" fails to couple to the limiting metric in accordance with equations (3.1.1).) Now let  $\tau$  be any topology on the space of rank  $(0, 2)$ -tensor fields with respect to which it might be claimed that  ${}_jT_{ab} \rightarrow T_{ab}$  as  $j \rightarrow \infty$ . Of course there are numerous topologies with respect to which this might be claimed; some may be physically appropriate and others may not. Luckily, we need not determine here which particular topology (if any) is in fact most appropriate. Instead we will only require that any relevant convergence must at least ensure the following condition for all  $p \in \mathcal{K}$ :

**Vacuum-point preservation:** If there exists a  $j_0 \in \mathbb{N}$  such that for all  $j > j_0$ ,  $T_{ab}|_p = \mathbf{0}$ , then  $T_{ab}|_p = \mathbf{0}$ .

The vacuum-point preservation condition should strike us as a reasonable restriction on any  $\tau$ -convergence in this context since it only precludes the sudden appearance of "new energy-momentum" at the limit that wasn't already present in the approach as  $j \rightarrow \infty$ . (Vacuum-point preservation would be obeyed, for instance, if we wanted to focus our discussion specifically on convergence in  $\mathcal{C}^2(\mathcal{K})$ .) Now from condition 1 of theorem 3.2, we know that for every  $j$ ,  ${}_jT_{ab} = {}_jG_{ab} = \mathbf{0}$  on  $\partial\mathcal{K}$ . Hence, if the limit is vacuum-point preserving we have that  $T_{ab}|_{\partial\mathcal{K}} = \mathbf{0}$ . But  $\mathcal{K}$  was an arbitrary sufficiently small neighborhood of  $\gamma$ , which means that  $T_{ab}$  will vanish on the boundary of *every* sufficiently small neighborhood of  $\gamma$ . Moreover, since  $g_{ab}$  is smooth, we know that  $T_{ab} = G_{ab}$  must be smooth. It follows from these two facts that  $T_{ab}$  must vanish in some neighborhood of  $\gamma$ . Hence, the only way to ensure that the energy-momentum tensors coupling to the sequences  $({}_jg_{ab})_{j \in \mathbb{N}}$  defined in theorem 3.2 converge in a vacuum-point preserving way without violating Einstein's field equations is by having them vanish around  $\gamma$ .

proofs, though the proofs can establish a kind of “approximation” to geodesic following in the case of extended bodies with matter-energy, when it comes to establishing genuine geodicity, both fall short, achieving such strict results (at best) only for inapposite or pathologically idealized special cases.

## 3.5 Towards a Geodesic Universality Thesis

In this chapter we have argued against the canonical view that the geodesic principle provides the dynamics of general relativity theory. Under this interpretation, the commonly endorsed belief that the principle can be derived either from Einstein’s original field equations or a distributional generalization of them must be rejected (even if we allow for further background assumptions about the kind of matter-energy that is supposed to follow such geodesics). By reviewing the three major classes of proof, we have seen that would-be geodesic following bodies are forced either **(i)** to meet unrealistically restrictive special-case conditions, **(ii)** to have no matter-energy at all (i.e. vanish), **(iii)** to violate Einstein’s field equations, or **(iv)** to be located on “paths” that don’t just fail to be geodetic but fail to exist in the spacetime manifold at all.

These results establish that models in general relativity of massive bodies precisely following some geodesic unavoidably qualify as examples of the *pathological idealizations* defined in chapter 2. Such pathologies associated with exact geodesic following reveal that the claim that “massive bodies follow geodesics in Einstein’s theory” cannot be accepted as a (precisely) faithful representation of the actual motion of bodies in general relativity. However, this does not mean that there is no place for the principle. We shall argue in the next chapter that it is possible to take advantage of the results discussed in section 3.4.2 to still develop a robust



### 3.5. TOWARDS A GEODESIC UNIVERSALITY THESIS

$\epsilon$ -faithful representation of dynamics in general relativity. Such  $\epsilon$ -faithful representations will appeal to the (pathologically) perfect geodesic cases, but instead of treating them as directly matched representations of relativistic dynamics, the pathological idealizations play the role of anchor models in the relevant  $\delta$  to  $\epsilon$  deductions. Further, taking advantage of the continuity of these  $\delta$  to  $\epsilon$  deductions, we will argue that as anchors, such pathological models can be used to understand the group behavior of entire classes of gravitating bodies in the form of a universality phenomena. Under this interpretation, though the geodesic principle can no longer play the role of a precise dynamical law governing the paths of massive bodies in Einstein's theory, it will still be recovered as a (less fundamental) *geodesic universality thesis* analogous to the kind of universality phenomena found in thermal systems during phase transitions.

## Chapter 4

# Geodicity and $\epsilon$ -Faithful Universality

In chapter 3, we saw that according to Einstein’s original conception of the general theory of relativity, the behavior of gravitating bodies was determined by two laws: The first (more fundamental) law consisted of his celebrated field equations describing how the geometry of spacetime is influenced by the flow of matter-energy. The second governing principle, referred to as the *geodesic principle*, then provides the “law of motion” for how a gravitating body will “surf the geometric field” as it moves through spacetime. According to this principle a gravitating body traces out the “straightest possible” or *geodesic* paths of the spacetime geometry. Not long after the theory’s initial introduction, it became apparent that the independent postulation of the geodesic principle to provide the theory’s law of motion was redundant. In contrast to classical electrodynamics and Newtonian gravitation, general relativity seemed special in that its dynamics providing principle could be derived directly from the field equations.

Though the motion of gravitating bodies is not logically independent of Einstein’s field equations, the geodesic principle canonically interpreted as providing a precise prescription for the dynamical evolution of massive bodies in general relativity does not follow from Ein-

stein’s field equations. To the contrary, in chapter 3 it was argued that under the canonical interpretation, *not only does the geodesic principle fail to follow from the field equations, but such exactly geodetic evolution would generically violate the field equations for non-vanishing massive bodies*. In short, under the canonical interpretation the two laws are not even consistent.

Despite this failure, the widespread “approximately geodetic” motion of free-fall bodies must not be denied. The nearly-geodetic evolution of gravitating bodies is frequently well confirmed within certain margins of error. Moreover, some of the most important confirmations of Einstein’s theory, including the classic recovery of the otherwise anomalous perihelion of Mercury, also appear to confirm the approximately geodetic motion of massive bodies. This abundance of apparent confirmation suggests that though the claim that massive bodies must exactly follow geodesics fails to cohere with Einstein’s theory, geodesic following may constitute some kind of *idealization* or *approximately correct* description of how generic massive bodies behave.

To understand the geodesic principle we must hence reconcile an apparent dilemma: On the one hand geodesic following appears illustrative as an ideal of the true motion of massive bodies. On the other hand the arguments against the canonical view in chapter 3 reveal that non-vanishing bodies that actually follow geodesics would be highly pathological with respect to the theory. So in order to gain knowledge about the paths of *actual* bodies, adopting such models as idealizations must come with an account of how models generically incompatible with the theory can still be epistemologically potent when it comes to understanding non-pathological targets.

In this chapter, we will reconcile this dilemma by taking advantage of the account of  $\epsilon$ -faithful representation presented in chapter 2. Further, by developing an analysis of the

#### 4.1. ANOTHER PLACE FOR THE PRINCIPLE?

concept of *universality phenomena*, we will identify an epistemic gain made possible through the explanatory role such wide scope representations can play by making use of pathological models to anchor the general behavior found across suitable classes of gravitating bodies. Because of the anchoring role played by geodesic models in such cases, we will propose that the geodesic principle be adopted as a *universality thesis* about the clustering of certain classes of gravitating bodies that exhibit *nearly*-geodetic motion.

We begin this investigation in section 4.1 by first taking a closer look at attitudes held by contemporary theoreticians towards the current role of the principle in general relativity. In section 4.2, we will proceed to develop an analysis of the general concept of *universality phenomena* to designate a certain kind of similarity of behavior exhibited across a wide class of (ostensibly diverse) systems of a particular theory. Using this analysis, in section 4.3, we will look at how the nearly geodetic behavior observed in numerous gravitational systems counts as such a clustering within appropriately close (topological) neighborhoods of geodesic anchor models. Finally, in section 4.4, we will explain why such pathological anchor models can be employed to characterize this clustering of the realistic models, taking advantage of the arguments presented in section 2.3 to show why it is not necessary to reify the pathological models in order to gain knowledge of actual physical systems.

### 4.1 Another place for the Principle?

The geodesic principle cannot accurately provide an account of the motion of massive bodies consistent with Einstein’s equations. Even in the case of “arbitrarily small” bodies, the principle fails to account precisely for the dynamics of general relativity theory. However, general relativity does have a well-posed initial value formulation. So, even if we were to purge

#### 4.1. ANOTHER PLACE FOR THE PRINCIPLE?

the principle from the theory, we would not leave it dynamically neutered. Geroch punctuates this spurious need for a geodesic principle (at least from a fundamental perspective) rather poetically as follows:

Matter in general relativity is normally described by fields subject to equations,... These equations have an initial-value formulation. So, the detailed motion of bodies (i.e., what every little part of the body is doing at every time) is determined within the theory. If one is lucky, one may be able to summarize this detailed motion by a slogan (e.g., “geodesic motion of the body as a whole”) in some limiting cases. If you can, more power to you; and if you can’t, well, this was only a diversion in any case. In other words, general relativity (as presently envisioned ...) has no need for - indeed no place for - any geodesic “postulate”.<sup>1</sup>

The well posed initial value formulation allows us (in principle) to evolve dynamically a set of initial conditions of the state of the world to the state at “future times” (i.e. fill in the future domain of dependence of a time-slice of some local region). Geroch emphasizes that this sort evolution of the of initial values is what (at least according to contemporary visions) more appropriately plays the “dynamical role” canonically thought to be filled by the geodesic principle. Strictly speaking, the geodesic principle is not necessary when it comes to satisfying this more “fundamental” dynamical need. Under the right circumstances and with enough information about the initial state we can (in principle) figure out what “every little part” of a body will be doing at every spacetime point.

So not only do the arguments reviewed in chapter 3 reveal that the geodesic principle should not be used to prescribe the precise dynamics of massive bodies in general relativity, strictly speaking we do not even need it to play this role. These two facts appear to be strong motivations for purging the geodesic principle from contemporary relativity theory. Nevertheless, there still remain some compelling reasons against such expurgation. Some of the reasons are simply pragmatic: We generally do not have access to all of the appropriately

---

<sup>1</sup>Personal correspondence 11/5/2010, reproduced with permission.

#### 4.1. ANOTHER PLACE FOR THE PRINCIPLE?

detailed initial data information required (e.g. it is difficult to ever attain precise information about the flow of matter-energy in the interior of Mercury). Moreover, even if we did have the appropriate initial data down to every detail, there are significant practical hurdles involved with finding solutions to initial value formulations not the least of which being that results will generically have to be numerical rather than analytic.

Highlighting these complications associated with relying exclusively on initial value formulations for relativistic dynamics, [Gralla & Wald \(2008, p2\)](#) have recently noted that “[a]lthough it is now possible to find solutions numerically in many cases of interest, it is difficult and cumbersome to do so, and one may overlook subtle effects and/or remain unenlightened about some basic general features of the solutions.” They go on to emphasize that “[t]herefore, it is of considerable interest to develop methods that yield approximate descriptions of motion in some cases of interest.”

Gralla and Wald’s point here is particularly germane to answering the question of whether there is a proper place for the geodesic principle: Even if we do have access to sufficient initial data, and we were able to bound the pragmatic hurdles (numerically if not analytically), having the precise details about what every piece of matter-energy is doing at every single point in spacetime runs the risk of obscuring what we should expect of the *general behavior* of bodies under gravitational influences. Perhaps counterintuitively, by taking a step back from the deluge of details that we might receive from an initial value result, searching for general approximations of the motion may yield information about what they refer to as the “basic general features of the solutions” that cannot be clearly identified through the more precise approach.

Being able to recognize the occurrence of certain basic features has epistemic as well as practical value in gaining a greater (though perhaps no longer “fundamental”) understanding

#### 4.1. ANOTHER PLACE FOR THE PRINCIPLE?

of the theory. However, it is just this sort of scientific knowledge that can be obfuscated if our attention were entirely focused on the precise evolution rather than more *general patterns of dynamical clustering*. Of course, by attending to such general patterns, there should be an inevitable loss in the way of accuracy. Claims about the kinds of patterns of dynamical clustering we should expect in general relativity theory will have to be approximate in a certain sense, but if we wish to understand not just the precise evolution, but also the broad characteristics of the paths of general gravitating bodies, certain “approximate descriptions” may indeed have more potency with respect to identifying such dynamical clustering patterns.

Remarkably, if we wish to understand the principle as playing this role of characterizing general behavior, then the limit operation proof of (Ehlers & Geroch, 2004), discussed in section 3.4.2, helps to provide a substantial justification for why we should expect such “approximate geodicity” of their paths. Recall, though theorem 3.2 was unable to establish perfect geodesic evolution of (actually massive) bodies, the interpretation of the theorem’s significance offered above was able to establish that “*appropriately small* massive bodies will follow timelike paths that are *almost* geodesic.” So while such a result fails to tell us about the paths with absolute precision, it does enable us to draw the broad inference that for suitable time scales, large classes of bodies can be expected to stick to paths that are “close” to being geodesics. If understood in context, being able to draw this kind of inference for such general classes of bodies without having to know about their exact constitution offers great opportunity for understanding about gravitation and gravitational dynamics despite the lack of attention to details at every level of precision, and this knowledge can be achieved even though the geodesic principle fails to provide the kind of fundamental dynamics-defining role expected by the canonical interpretation.

## 4.2. UNIVERSALITY IN PHYSICS

The kind of near-geodesic clustering of massive bodies, which we will argue can be explained by results such as (Ehlers & Geroch, 2004), is well confirmed by our experimental observations.<sup>2</sup> Planetary bodies of relatively small size and gravitational effect compared to the sun actually exhibit nearly geodesic motion. In particular, the Mercury confirmation shows that this clustering can be confirmed in relativistic regimes. Though a precise enough experiment should reveal divergence from perfect geodicity, as a thesis about the near-geodesic clustering of freely gravitating bodies in general, the kind of subtle wobbling we might expect given the discussion in section 3.3 need not count as disconfirmation. Instead, a vast number of examples from Newton’s apple to gravitation on astronomical scales can constitute opportunities to confirm such general near-geodesic clustering in the form of a thesis about the kind of universality phenomena that we analyze in the next section.

## 4.2 Universality in Physics

The suggestion considered above that the geodesic principle might be reinterpreted as a characterization of the general patterns of behavior of (small) gravitating bodies (despite significant possible variations in details of how the bodies are constituted or the type of external gravitational field they might be exposed to) is analogous to a prominent classification of certain phenomena studied in other fields of physics. Referred to as *universality phenomena*, such clustering patterns across multiple systems were originally used to characterize the similarities in behavior exhibited by thermal systems during phase transitions and near criticality. Kadanoff (2000, p225), often identified as one of the first to apply this concept

---

<sup>2</sup>In fact, as long as we do not expect perfect geodicity, results from more sophisticated “0<sup>th</sup>-order type” proofs (e.g. those in (Gralla & Wald, 2008, Pound, 2010)) can be used to identify the appropriate regime scales for which we might expect such observed clustering.



## 4.2. UNIVERSALITY IN PHYSICS

in its contemporary sense in physics, has defined the concept of ‘universality’ as applying to patterns in which “[m]any physically different systems show the same behavior.” In the study of critical behavior, for instance, the phenomena is identified when numerous systems seem to cluster into what are called *universality classes*, despite possible vast disparities in the fundamental details characterizing different members of a class.

In the next two sections we will take a closer look at the concept of ‘universality phenomena.’ We begin with a discussion in section 4.2.1 of the most well known example of universality observed in thermal systems undergoing phase transitions. In section 4.2.2 we will proceed to offer a general analysis of universality phenomena in the context of particular theories.

### 4.2.1 The Paradigm Case: Universality in Phase Transitions

The notion of a universality phenomenon was initially appealed to in order to characterize a remarkable clustering in the behavior of thermal systems undergoing phase transitions, particularly the behavior of systems in the vicinity of a thermodynamic state called the “critical point.” In thermodynamics the state of a system can be characterized by the three state variables pressure ( $P$ ), temperature ( $T$ ), and density ( $\rho$ ). In figure 4.2.1 we see a phase diagram of some generic material projected onto the pressure-temperature plane. As discussed in chapters 1 and 2, according to the thermodynamic study of phase transitions, when the state of a system is kept below the particular “critical point” values ( $P_c, T_c, \rho_c$ ) associated with the substance, phase transition boundaries correspond to discrete changes in the system. These boundaries are signified in our diagram by the thick black lines. If, however, a system is allowed to exceed its critical values, there exist paths available to the

## 4.2. UNIVERSALITY IN PHYSICS

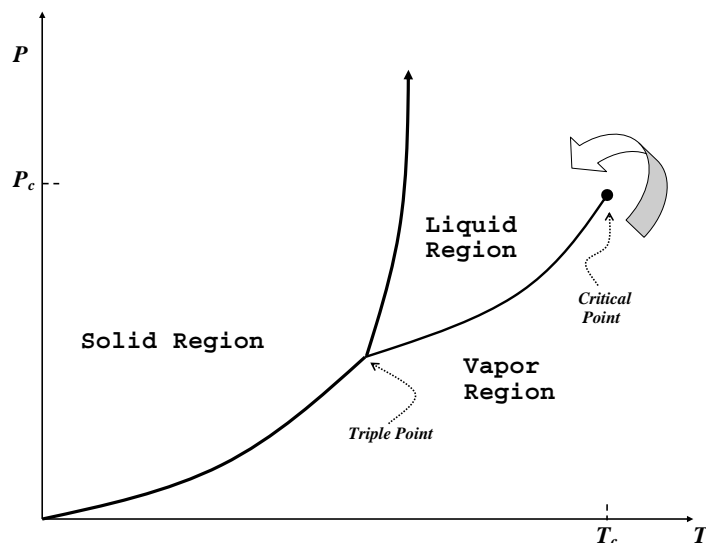


Figure 4.2.1: *Phase diagram of a generic material at fixed density.*

system allowing it to change from vapor to liquid (or back) without undergoing such discrete changes. These paths involve avoiding the vapor-liquid boundary line by navigating around the critical point as depicted by the broad arrow in figure 4.2.1.

There exists a remarkable uniformity in the behavior of different systems near the critical point. One such uniformity is depicted in figure 4.2.2. In this figure we see a plot of data recovered by Guggenheim (1945) in a temperature-density graph of the thermodynamic states at which various fluids transition from a liquid or vapor state to a “two phase” liquid-vapor coexistence region. Systems in states located in this latter region can be in liquid or vapor phases and (according to thermodynamics) will maintain constant temperature as the density of the system changes. An important feature exhibited in figure 4.2.2 is that (after rescaling for the  $\rho_c$  and  $T_c$  of the respective molecules) the transition data of each of the distinct substances near criticality appears to be well fit by a *single curve* referred to as the *coexistence curve*. This similarity in the coexistence curves best fitting diverse molecular substances can be characterized by a particular value  $\beta$  referred to as the *critical exponent*

## 4.2. UNIVERSALITY IN PHYSICS

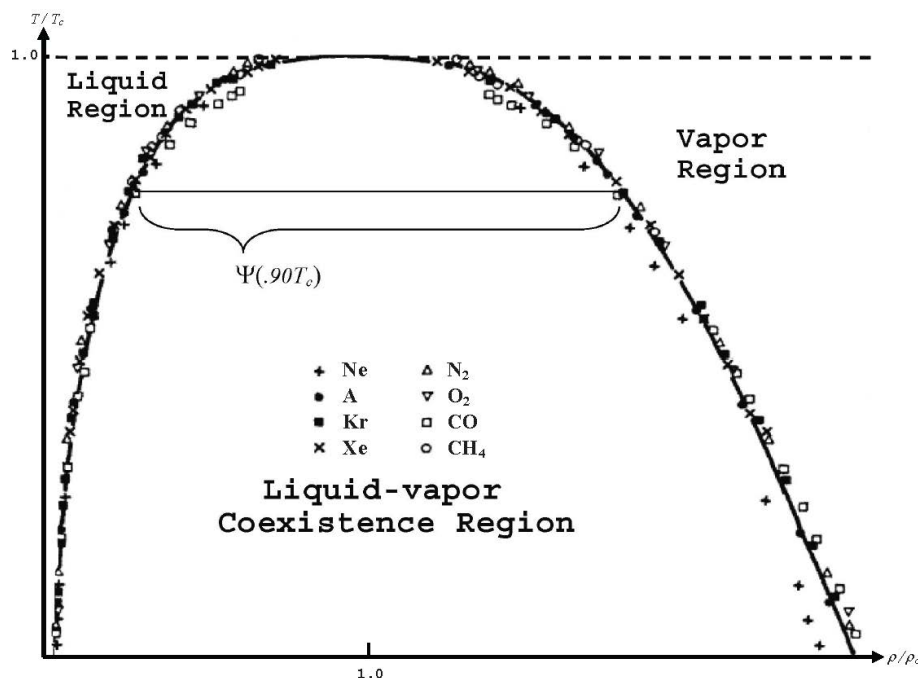


Figure 4.2.2: Adapted plot of (*Guggenheim, 1945*) data rescaled for criticality.

found in the following relation:

$$\Psi(T) \propto \left| \frac{T - T_c}{T_c} \right|^\beta \quad (4.2.1)$$

where the parameter  $\Psi(T)$ , called the *order parameter* tells us the width of the coexistence curve at a particular temperature value  $T$ . As depicted in figure 4.2.2, as  $T$  gets closer and closer to the critical temperature  $T_c$  from below, this width drops down eventually vanishing at criticality. We can think of the critical exponent  $\beta$  as telling us about how rapidly such a vanishing occurs. As confirmed by the above data, this number turns out to be similar (in the neighborhood of  $\beta \simeq .33$ ) for vastly different fluid substances.<sup>3</sup>

What is fascinating about examples such as this is not the universal (or “nearly” univer-

---

<sup>3</sup>This similarity in the value of the critical exponent exists not only for thermal fluid systems, but also in describing the behavior of ferromagnetic systems in the neighborhood of a thermal state that can be analogously characterized as the critical point.

## 4.2. UNIVERSALITY IN PHYSICS

sal) regularity in physical systems. That uniform reliable regularities (viz “universal laws”) can be found to apply to numerous physical systems (though remarkable) is nothing new. The interesting part is that such uniform reliable behavior occurs *despite the fact that at least at one level of description the systems are so incredibly dissimilar*. From a level of description thought to be perhaps more “fundamental” than the gross state variables ( $P$ ,  $T$ , and  $\rho$ ) used to characterize thermodynamic systems, the various substances exhibiting similar critical exponent values have quite diverse descriptions: At the quantum mechanical level, for instance, the state vectors or density matrices representing the respective quantum mixtures will be incredibly distinct (e.g. close to orthogonal). Moreover, we need not go down to a quantum level of description to recognize the vast diversity. From a chemical perspective monotonic neon is different from a diatomic oxygen molecule, or an asymmetrical carbon monoxide molecule. We might hence expect surprise from a physicist or chemist, analyzing the behavior of these systems at these respective levels of theorizing, over the fact that despite such vast differences in the ostensibly pertinent details at these levels, the substances still share this observed similarity. This similarity despite such (speciously relevant) differences is what distinguishes the behavior across thermal systems as a kind of *universality phenomenon*. In the next section we will begin a more explicit analysis of the concept’s general application in physics.

Before proceeding, it is worth noting that though the usage of the term originated in the study of thermal systems, universality has now been identified in a multitude of other domains. Over the past decade, Robert Batterman has argued convincingly in the philosophical literature that “while most discussions of universality and its explanation take place in the context of thermodynamics and statistical mechanics,... universal behavior is really ubiquitous in science” (Batterman, 2002). A (far from comprehensive) list of vindicating

## 4.2. UNIVERSALITY IN PHYSICS

examples includes the clustering behavior found in contexts including non-thermal criticality patterns exhibited in avalanche and earthquake modeling (Kadanoff *et al.*, 1989, Lise & Paczuski, 2001), extinction modeling in population genetics (Sole & Manrubia, 1996), and belief propagation modeling in multi-agent networks (Glinton *et al.*, 2007, 2010). Batterman has discussed many examples of universality phenomena distinct from criticality phenomena, including patterns in rainbow formation, semi-classical approximation, and drop breaking (Batterman, 2002, 2005, 2006, 2009). Numerous non-criticality examples of universality have also been discovered in contexts such as the study of chaotic systems exhibiting “universal ratios” in period doubling (Feigenbaum, 1978, Hu & Mao, 1982), or the clustering similarities in models of cold dark matter halos found in astronomical observations (Navarro *et al.*, 1997, 2004), to name a couple. In the next section we will begin a more explicit analysis of the concept’s general application in physics.

### 4.2.2 The Same but Different: Analyzing Universality

The term *universality* is generally used in physics to describe cases in which broad similarities are exhibited by classes of physical systems despite possibly significant variations according to apparently “more fundamental” representations of the systems. Recall, Kadanoff (2000, p225) describes the term most generally as applying to those patterns in which “[m]any physically different systems show the same behavior.” Batterman (2002, p4) explains that the “essence of universality” can be found when “many systems exhibit similar or identical behavior despite the fact that they are, at base, physically quite distinct,” and Berry (1987) has described it as the “way in which physicists denote identical behavior in different systems.” Characterizations such as these reveal that the concept hinges on the satisfaction of the two seemingly competing conditions of displaying a particular *similarity* despite other (evidently

## 4.2. UNIVERSALITY IN PHYSICS

irrelevant) *differences* in the systems at some level of description. To make this conceptual dependency explicit, let us propose the following analysis of *universality phenomena*.

**(UP):** A class  $X_{\mathcal{T}}$  of models of physical systems in a theoretical context  $\mathcal{T}$  will be said to exhibit a *universality phenomenon* whenever the class can simultaneously meet the following two conditions:

- (Sim) There exists a robust similarity in some observable behavior across the physical systems modeled by members of  $X_{\mathcal{T}}$ .
- (Var) This similarity in the behavior of members modeled in  $X_{\mathcal{T}}$  is stable under robust variations of their state descriptions according to context  $\mathcal{T}$ .

The first thing to specify is what counts as a “class of models of physical systems in a theoretical context.” In order to avoid complications associated with multiple (possibly not entirely equivalent) formulations of a full physical theory, **(UP)** is best analyzed in terms of the more restrictive notion of a theoretical context  $\mathcal{T}$  which identifies within a given theory a particular formulation and variety of studied phenomena. Examples of different theoretical contexts in classical mechanics include the Hamiltonian versus the Lagrangian formulations, or in quantum mechanics we might distinguish between wave mechanics and operator mechanics.<sup>4</sup> A theoretical context may also restrict the phenomena considered by the total theory. For example, *source free* classical electrodynamics might be considered a distinct theoretical context within the full theory of classical electrodynamics which also models the effects of sources. In some cases it is possible for a theoretical context  $\mathcal{T}$  to specify an entire theory uniquely; in other cases, a specification in terms of (potentially non-equivalent) formulations and specific phenomena types may be appropriate.

---

<sup>4</sup>Note, in both dichotomies there exist occasional circumstances or conditions such that the alternate formulations can cease to be equivalent.

## 4.2. UNIVERSALITY IN PHYSICS

Given a particular theoretical context  $\mathcal{T}$  of a universality phenomena, the expert will typically be able to identify pertinent state descriptions “according to context  $\mathcal{T}$ .” For example, in classical electromagnetism the relevant state description may come in the form of fields specifying the flow of the source charges and the electromagnetic field values throughout a spacetime; in general relativity the metric and energy-momentum tensors might play this role; in thermodynamics, state descriptions may be parametrized by  $P$ ,  $T$ , and  $\rho$  (or perhaps  $V$  and  $N$ ), whereas in quantum statistical mechanics one may use density operators.

Satisfaction of (Sim) is primarily an empirical question. In order to claim that something universality-like is occurring, there must be an evident similarity in the class of systems exhibiting the phenomenon. This evident similarity need not be (directly) in terms of any of the state descriptions used to characterize elements of  $X_{\mathcal{T}}$ . So for the paradigm example of the universality of phase transitions, (Sim) is satisfied once physicists recover sufficient empirical data of the kind depicted in figure 4.2.2. The robust similarity of (Sim) can be quantified in terms of the remarkable closeness of the critical exponents of these various systems even though the critical exponent parameter  $\beta$  may not necessarily be put in terms of the state quantities of  $\mathcal{T}$  (e.g. chemistry or statistical mechanics). In the context of our discussion of section 1.3, this latter step of identifying the remarkable closeness exemplifies the fact that determining whether or not (Sim) has been satisfied is not strictly a matter of observing the data. In order to identify the relevant robust similarity necessary for (Sim), phenomenal patterns must be abstracted from this data in a way that permits us to identify how the phenomenal behavior detected by the observed data is similar. In the thermal case, this (ampliative) data to phenomena move occurs when the coexistence curve is fit to the data after rescaling for criticality.

## 4.2. UNIVERSALITY IN PHYSICS

Satisfaction of (Var) will depend primarily on the size and most importantly the diversity of the models in class  $X_{\mathcal{T}}$ . The larger and more varied the members of class  $X_{\mathcal{T}}$  with respect to the relevant state descriptions of  $\mathcal{T}$ , the more “stable under variations.” If  $X_{\mathcal{T}}$  is suitably rich with diverse members, then a member  $x \in X_{\mathcal{T}}$  may be “mapped” to a rich variety of other members of  $X_{\mathcal{T}}$  while still maintaining the very similarity shared by all members of  $X_{\mathcal{T}}$  that allowed the class to satisfy (Sim). In the paradigm example of thermal universality, (Var) is satisfied by the fact that at the chemical or the statistical mechanics levels of description, the members in our class sharing this similar critical behavior are so diverse.<sup>5</sup>

Though satisfaction of the (UP) conditions will be evaluated in the above terms, note that the central concepts of *robust variation* and *robust similarity* on which (Var) and (Sim) respectively depend are not binary. We can think of some universality phenomena as “more robust” than other instances, in terms of both the “degree” of similarity displayed and the “degree” of variations that the systems can withstand while still exhibiting such similar behavior. The greater the robustness of the pertinent similarity in behavior across the class of systems and the more variation (with respect to the irrelevant differences in the  $\mathcal{T}$ -state descriptions) found within the class, the more robust the universality is.<sup>6</sup> This dependence on non-binary conditions means the concept of universality may be subject to vagueness challenges in some cases. While certain examples, such as thermal criticality behavior and, as we shall argue, the clustering behavior of free-fall massive bodies around geodesic paths may be identified as determinant cases of universality, penumbral cases where it is unclear

---

<sup>5</sup>As already discussed, the substances depicted in figure 4.2.2 are quite varied in their molecular structure, and at the quantum level we might quantify this point with an appeal to vanishingly small transition probabilities.

<sup>6</sup>Often this can be rigorously assessed by an appropriately natural norm, metric, topology, etc. defined on the space of state descriptions of  $\mathcal{T}$ . E.g. we might use some integration norm to quantify the difference between two (scalar) fields found in  $X_{\mathcal{T}}$ . Of course, the choice of appropriate norm, topology, etc. identifying differences in the members of  $X_{\mathcal{T}}$  will inevitably be dependent on the context  $\mathcal{T}$ .



whether a candidate universality class is sufficiently similar and robust under variations may exist.

## 4.3 The Geodesic Universality Thesis

In this section we reconsider the case of near-geodesic clustering observed in nature in terms of the **(UP)** analysis. In 4.3.1 we will examine why such clustering qualifies as an example of a universality phenomenon. In 4.3.2 we will then identify how the limit operation result of Ehlers and Geroch offers what we will identify as a *universality explanation* of this clustering.

### 4.3.1 The Similarity and Diversity of Geodesic Universality

Consider a sequence of classes  $(X_{\mathcal{GR}}^\epsilon)_{\epsilon \in (0,s)}$  indexed by some sufficiently small error parameter  $\epsilon \in (0, s)$ . For fixed  $\epsilon$ , the class  $X_{\mathcal{GR}}^\epsilon$  consists of (local) solutions to Einstein’s field equations (3.1.1). Each member of  $X_{\mathcal{GR}}^\epsilon$  models some massive body whose spacetime path “comes close to following” a (timelike) curve  $\gamma$  that is “close to actually being a geodesic” (where these two senses of closeness are parametrized by respective functions monotonically vanishing with the smallness of  $\epsilon$ ).<sup>7</sup> With the **(UP)** analysis in hand, for a given degree of “ $\epsilon$ -closeness” we can now ask if such a class  $X_{\mathcal{GR}}^\epsilon$  satisfies the (Sim) and (Var) conditions in the context of general relativity theory purged of the canonical commitment to geodesic dynamics argued against in chapter 3.

The satisfaction of (Sim) is an empirical matter apparently well confirmed by centuries

---

<sup>7</sup>For example, the “closeness to geodcity” of  $\gamma$  can be put in terms of the vector field values  $\xi^a \nabla_a \xi^b$  along the points of  $\gamma$  where the  $\nabla$  are the unique connections of the respective spacetime metrics  $g_{ab}$  of the (local) solutions, and  $\xi^a$  is the (timelike) tangent vector field of  $\gamma$ . Assigning numerical quantification to our  $\epsilon$  values can be achieved with the selection of a chart, allowing us e.g. to bound the absolute interval magnitudes of the (spacelike)  $\xi^a \nabla_a \xi^b$  by an appropriate monotonic function of the  $\epsilon$  values. A similar quantification of “closeness of following  $\gamma$ ” can be evaluated in terms of the maximum (spacelike geodesic) distance of the body from  $\gamma$  for points along the curve.

### 4.3. THE GEODESIC UNIVERSALITY THESIS

of astronomical data recovered from cases in which a relatively small body (a planet, moon, satellite, comet, or even a star) travels under the influence of a much stronger gravitational source. Examples involving non-negligible relativistic effects (like the Mercury confirmation) are of particular importance, but even terrestrial cases including Galileo and leaning towers or other (nearly) free-fall examples in determinately Newtonian regimes can count as confirming instances for certain  $\epsilon$ -closeness values. Since observational precision is inevitably bounded, in many of these cases it is often claimed that the satellite, moon, planet, etc. indeed “follows a geodesic,” despite the results of chapter 3. In such instances, a body is actually observed to come “close enough” to following a geodesic to warrant such equivocation. These instances hence confirm membership in a class  $X_{\mathcal{GR}}^\epsilon$  for some  $\epsilon$  threshold below the level of experimental precision or attention.

In order to appreciate the satisfaction of (Var), we must consider the relevant theoretical context of general relativity theory. State descriptions of physical systems according to the theory come in the form of the tensor fields  $T_{ab}$  and  $g_{ab}$ , related by the equations (3.1.1). Assuming we only consider (local) solutions to Einstein’s equations, there exist six independent field components describing  $g_{ab}$  and so the matter-energy flow  $T_{ab}$ . In other words, from a fundamentals of relativity theory perspective, there are six physical degrees of freedom to how these bodies are described at each spacetime point.

Given the wealth of evident confirming instances falling under a class  $X_{\mathcal{GR}}^\epsilon$  with suitable  $\epsilon$ , there will be significant variation in terms of these degrees (even after rescaling) once we consider the significant differences in the density, shape and flow of the matter-energy of a planet, versus a satellite, asteroid, anvil, etc. In these “fundamental state description” terms, the diversity of the bodies in a given class  $X_{\mathcal{GR}}^\epsilon$  will be quite significant (again even after rescaling with respect to the background). Despite this diversity, such bodies still satisfy the

### 4.3. THE GEODESIC UNIVERSALITY THESIS

defining requirement of  $\epsilon$ -closeness to following a geodesic. It is with respect to this diversity in these degrees of freedom (of the energy-momenta/gravitational influences of the “near-geodesic following bodies” of members in  $X_{\mathcal{GR}}^\epsilon$ ) that a “robust stability under variations” can be established in accordance with (Var).

So, according to our **(UP)** analysis, such near-geodesic clustering observed in nature constitutes a *geodesic universality phenomenon*. However, meeting the conditions of the analysis depends entirely on the truth of the above made *empirical* claims about the existence of bodies well modeled by some member of the respective  $X_{\mathcal{GR}}^\epsilon$  classes for a suitable range of  $\epsilon$  values, and that the bodies in each class are so fantastically diverse from the perspective of their  $T_{ab}$  ( $g_{ab}$ ) fields. In the next section we will turn to the more *theoretical* question of understanding how such geodesic universality is possible in general relativity, by considering the properties of the classes  $(X_{\mathcal{GR}}^\epsilon)_{\epsilon \in (0,s)}$  in terms of the important geodesic result of [Ehlers & Geroch \(2004\)](#).

#### 4.3.2 Explaining Geodesic Universality

We have now formulated the geodesic universality thesis in the context of general relativity as an empirically contingent claim about classes of the form  $X_{\mathcal{GR}}^\epsilon$  whose members model a physical system such that the path of some body counts as  $\epsilon$ -close to being geodesic without violating Einstein’s field equations. We have also given a plausibility argument suggesting why a good deal of observational data already obtained by experimentalists confirms this empirical hypothesis. Moreover, given such confirmation and the diversity of the energy-momenta of the respective bodies, membership in some  $X_{\mathcal{GR}}^\epsilon$  will be sufficiently stable under significant variations of the fundamental state descriptions of the theory to satisfy (Var). A remaining theoretical question must now be answered: *How can the systems exhibiting this*

### 4.3. THE GEODESIC UNIVERSALITY THESIS

*universality phenomenon behave so similarly while being so different at the level of theoretical description fundamental to general relativity?*

Geodesic universality can be explained by appealing to “limit proofs” of the geodesic principle discussed in section 3.4. It was argued there that [Ehlers & Geroch \(2004\)](#) are able to deduce the “approximate geodesic motion” of gravitating bodies with relatively small volume and gravitational influence, by considering sequences of energy-momentum tensor fields with positive mass of the form  $(T_{(i,j)}^{ab})_{i,j \in \mathbb{N}}$ , referred to as “EG-particles.” The spatial extent and gravitational influence of these EG-particles can be made arbitrarily small by picking sufficiently large  $i$  and  $j$  values respectively. The theorem of ([Ehlers & Geroch, 2004](#)) entails that if for a given curve  $\gamma$  there exists such an EG-particle sequence, then by picking a large enough  $j$ ,  $\gamma$  comes arbitrarily close to becoming a geodesic in a spacetime containing the  $T_{(i,j)}^{ab}$  instantiated matter-energy.

Specifically, let  $(g_{(i,j)}^{ab})_{i,j \in \mathbb{N}}$  be the sequence of metrics that couple to these  $(T_{(i,j)}^{ab})_{i,j \in \mathbb{N}}$  according to (3.1.1) in arbitrarily small neighborhoods  $(\mathcal{K}_i)_{i \in \mathbb{N}}$  of  $\gamma$ , containing the support of the respective  $T_{(i,j)}^{ab}$ . Then if for each  $i$ , as  $j \rightarrow \infty$  the  $g_{(i,j)}^{ab}$  approach a “limit metric”  $g_{ab}$  in the  $\mathcal{C}^1(\mathcal{K}_i)$  topology, which keeps track of differences in the metrics and their unique connections, then the curve  $\gamma$  approaches geodeticity as  $j \rightarrow \infty$ .

We can understand the impact of the theorem for our universality classes  $(X_{\mathcal{GR}}^\epsilon)_{\epsilon \in (0,s)}$  in terms of the deductive continuity of the resulting  $\epsilon$ - $\delta$  relationship, as defined in section 2.2. The limiting behavior of the Ehlers-Geroch result establishes exactly this kind of relationship between (on the  $\epsilon$ -end) how “nearly-geodetic” we want the curve around which the body travels to be, and (on the  $\delta$ -end) how much we need to bound the gravitational effects of the body on the background spacetime.<sup>8</sup> That is to say, the Ehlers-Geroch limit result can be

---

<sup>8</sup>See also the limit results in ([Gralla & Wald, 2008](#), §§3-5).

### 4.3. THE GEODESIC UNIVERSALITY THESIS

thought of as telling us that “for every degree of  $\epsilon$ -closeness to geodicty we want the bodies’ path to be, there exists a  $\delta$ -bound on the gravitational effect of the body that will keep the path at least that close to geodicty.” The important thing to observe about why this  $\epsilon$ - $\delta$  interplay works is that though the limiting relationship *does* require imposing a  $\delta$ -bound (in terms of the respective  $\mathcal{C}^1$  topologies) on the perturbative effects of the body, it does not impose any *specific constraints on the details* of how the matter-energy of the body flows within in the ( $\epsilon$ -close) spatial neighborhood ( $\mathcal{K}$ ) of the curve, nor how the metric it couples to specifically behaves. So though the metric is “bounded” within a certain  $\delta$ -neighborhood of the limit metric, the particular details of the tensor values, the corresponding connection, and especially the curvature have considerable room for variation so long as they stay “bounded in that neighborhood.”

This relationship established by the Ehlers-Geroch theorem hence gives us a kind of *details-free* way of understanding the diverse populations of our respective universality classes  $(X_{\mathcal{GR}}^\epsilon)_{\epsilon \in (0,s)}$ . In effect the Ehlers-Geroch limiting relationship highlights that for each  $X_{\mathcal{GR}}^\epsilon$  class, there exists a particular  $\delta$ -bound around a limit metric with some geodesic anchor  $\gamma$  such that any body coupling to a metric that stays within that bound (in addition to remaining spatially close enough to  $\gamma$ ) will satisfy the relevant  $\epsilon$ -closeness part of the requirements for membership in  $X_{\mathcal{GR}}^\epsilon$ . This deduction enables  $\epsilon$ -fidelity with respect to knowledge of the nearly geodetic behavior of particular  $\delta$ -close members in a class  $X_{\mathcal{GR}}^\epsilon$ , but it also allows us to speak generally about entire subsets of  $X_{\mathcal{GR}}^\epsilon$  whose membership is entailed by their  $\delta$ -proximity. But as we just emphasized, *falling under this  $\delta$ -bound does not impose specific constraints on the detailed values of the energy-momenta or metric fields*. So, membership in the universality class  $X_{\mathcal{GR}}^\epsilon$  is possible as long as the body is a massive solution to Einstein’s equations, and its gravitational effect and extent are bounded in the right way.

#### 4.4. EXPLANATION WITHOUT REIFICATION

Beyond these requirements the specific details concerning “what the gravitational effect does below those bounds” are irrelevant. Hence, the limit behavior established by the Ehlers-Geroch theorem explains how the universality of this  $\epsilon$ -clustering near geodesic anchors is possible despite significant differences in the energy-momenta of our near-geodesic following bodies: So long as the bodies’ gravitational influences are bounded in the right way their (positive) matter-energy can vary as much as we like under those bounds.

This explanation is an example of the special kind of knowledge made possible by the  $\epsilon$ -fidelity of our representation that cannot even be broached in the context of narrow scope matching accounts.<sup>9</sup> By carefully investigating the right topological proximity, in particular the  $\delta$ -bounds delineated by the  $\mathcal{C}^1$ -proximity to some “geodesic anchor model,” we can understand how it is possible to have so much diversity in a universality class  $X_{\mathcal{GR}}^\epsilon$  despite potentially significant variation in their tensor field state descriptions. In the next section, we will return to the arguments of section 2.3, where we discussed the role of such pathological anchor models in establishing this  $\epsilon$ -fidelity.

## 4.4 Explanation without Reification

Before closing there remains a potential challenge concerning how we can endorse any kind of geodesic “idealization” thesis if the actual geodesic motion of massive bodies is incompatible with Einstein’s theory. Recall, while explaining how the classes  $X_{\mathcal{GR}}^\epsilon$  whose respective members are “ $\epsilon$ -close” to geodesic following models could be so diverse, we needed to take the “geodesic limit” of the metrics  $(g_{ab})_{i,j \in \mathbb{N}}$  coupling to the EG-particles  $(T_{ab})_{i,j \in \mathbb{N}}$  in accordance with the equations (3.1.1). By taking such a “geodesic limit” in order to iden-

---

<sup>9</sup>See (Batterman, 2010) for a parallel discussion of how narrow scope matching accounts are incapable of mathematically representing universality phenomena anchored by pathological idealizations.

#### 4.4. EXPLANATION WITHOUT REIFICATION

tify the diversity of the kinds of solutions populating our  $X_{\mathcal{GR}}^\epsilon$  classes, haven't we made an "essential" appeal to the kind of pathological models precluded by Einstein's field equations?

The answer to this challenge hinges on the importance emphasized in section 2.3.2 of keeping track of the types of topological closeness we are appealing to in our (continuous)  $\epsilon$ - $\delta$  deductions. Recall, though the  $(g_{ab})_{i,j \in \mathbb{N}}$  converge to a well defined "geodesic limit" (in the  $\mathcal{C}^1$  topologies) the coupled energy-momentum tensors  $(T_{ab})_{i,j \in \mathbb{N}}$  may not. Moreover, even if they do converge in a physically salient and independently well defined way, "at the limit" they must either fail to obey the field equations (3.1.1) or vanish. The reason why this is not a problem goes back to the fact, emphasized in section 2.2.1, that the property mappings to the respective P and Q property spaces need not preserve *every* property and closeness relation to be found in the abstracted space of models S. In the current example, for fixed  $\mathcal{K}$ , the P-properties and relations are restricted to the metrics, their uniquely determined connections, and their  $\mathcal{C}^1(\mathcal{K})$  proximities. Consequently, the curvature properties associated with these metrics (and so which energy-momentum tensors they couple with according to (3.1.1)) do not make any difference in the  $\delta$ -proximity used in the  $\delta$  to  $\epsilon$  deductions. Similarly, the relevant Q-properties only track the "closeness" to geodicity of  $\gamma$  (as outlined in note 7). Neither of these property spaces is equipped with a less finely grained topology  $\tau_{Field\ Equations?}$  that (analogously to the course topologies  $\tau_{Analytic?}$  and  $\tau_{2?}$  of section 2.3.2) discretely sorts the  $(g_{ab}, T_{ab})$  pairs strictly on the basis of whether or not they obey the field equations (3.1.1).

In contrast, falling within an appropriate  $\delta$ -neighborhood of the geodesic limit in a  $\mathcal{C}^1$  topology was essential to our explanation of geodesic universality. If we allow in our abstraction of the space of models (S) to be broad enough (e.g. we allow it to include  $(g_{ab}, T_{ab})$  pairs that may not obey 3.1.1), then there can exist "geodesic models" to "anchor" these  $\delta$ -neighborhoods in that (at the Q-properties end) the relevant curve  $\gamma$  is actually a geodesic

#### 4.4. EXPLANATION WITHOUT REIFICATION

of the metric, and (at the P-properties end) the metric and connection in the  $\mathcal{K}$  spacetime neighborhoods around  $\gamma$  are identical to the relevant “background” metric and connection. Whether such field equations violating models are included or not,<sup>10</sup> as argued for at length in section 2.3, the role played by *geodesic anchor models* does not require us to reify the idealization, matching it with any physical system. Even though there are significant complications associated with what happens *at* the geodesic limit, the  $\delta$  to  $\epsilon$  deductions are based on well defined topologies describing the *approach* to the limiting anchor properties in the respective property spaces, and the behavior of the models in  $X_{\mathcal{GR}}^\epsilon$ , which are “close but not identical to” a “geodesic anchor model” exhibiting these properties, all still obey Einstein’s theory. Hence, they can still be well matched with physical targets that obey the field equations. The only thing such geodesic anchor models do in our ( $\epsilon$ -faithful) representation is exemplify the appropriate *anchor properties* of respective (topological) neighborhoods, but they need not be matched with any physical target (field equations violating or not) in order to anchor these properties for the elements of the respective  $X_{\mathcal{GR}}^\epsilon$  classes to cluster around them. Hence, using these models as anchors to identify the points around which the *actual solutions* to Einstein’s equations cluster does not require that the anchors themselves be admitted in  $X_{\mathcal{GR}}^\epsilon$ .

Universality phenomena (understood through  $\epsilon$ -faithful representations of sets of mathematical models) are about the *group behavior of classes of  $X_{\mathcal{T}}$*  not *individual systems*. For “narrow scope” representations requiring perfect matching, severe pathologies can be detrimental because they render *the sole idealized model* theoretically inapposite. In contrast, when we are representing universality with wide scope matching, the existence of a patho-

---

<sup>10</sup>Note, such anchor models need not necessarily violate the equations (3.1.1). Models with vanishing energy-momentum in some neighborhood containing  $\gamma$  can also suffice as anchors in that such solutions can possess each of these anchor properties exactly.



#### 4.4. EXPLANATION WITHOUT REIFICATION

logically idealized model “close to but excluded from” a universality class  $X_{\mathcal{T}}$  or sequence of classes  $(X_{\mathcal{T}}^{\epsilon})_{\epsilon \in I}$  need not entail that members of the class(es) are likewise poorly behaved. Moreover, if a topological clustering “near” to an idealized model has physical significance (as with the  $\mathcal{C}^1$  topologies), such proximity enables  $\epsilon$ -fidelity, providing knowledge about systems that can match members of the well-behaved classes without molesting the models’ admissibility according to the laws of  $\mathcal{T}$ .

This is precisely what occurs in the case of geodesic universality. Members of a class  $X_{\mathcal{GR}}^{\epsilon}$  are able to take advantage of their closeness to the pathological geodesic anchor models without “contracting” any of the problems occurring *at* the actual geodesic limits. Moreover, we were able to *explain* such  $\epsilon$ -closeness by appealing to what we characterized as the “specific details irrelevant”  $\delta$ -closeness in the  $\mathcal{C}^1$  topologies. The  $\epsilon$ -fidelity of our representation hence comes in two forms: The  $\delta$  to  $\epsilon$  deductions allow us to know that a particular  $\delta$ -close model will follow “almost” geodesics, but it also allows us to gain knowledge about general dynamical clustering in Einstein’s theory. These deductions allowed us to understand why entire classes  $X_{\mathcal{GR}}^{\epsilon}$  of  $\epsilon$ -close to geodesic models are able to exhibit so much diversity and still share this significant similarity. The  $\epsilon$ -fidelity of our representation, hence, plays an important role in our understanding of the phenomena of geodesic universality.

# Chapter 5

## Conclusion

The primary question we have now answered is how mathematical representations enable knowledge of actual systems in the physical world. Our solution follows once we account for the imprecisions inevitable in the process of developing mathematical models of physical targets. We argued that this can be done with the condition we called  $\epsilon$ -*fidelity*, where a mathematical representation is  $\epsilon$ -faithful when it enables knowledge about physical targets not with absolute precision, but within suitable “ $\epsilon$ ” margins of error. In chapter 2, we demonstrated that such  $\epsilon$ -fidelity can be established through the use of  $\delta$  to  $\epsilon$  deductions that take advantage of “wide scope” potential relations abstracted from different physical systems of similar type. Such deductions allow us to ( $\epsilon$ -faithfully) gain knowledge about any physical target in some  $\delta$ -neighborhood of meeting the conditions required for the deduction, without demanding that they match such conditions at every level of precision. By relaxing the conditions presupposed in order to match a physical target, we argued that the knowledge gained by the resulting  $\epsilon$ -faithful mathematical representations can avoid the epistemologically untenable requirement that the particular physical system must meet some given constraint with absolute precision.

Multiple epistemological benefits follow from this relaxation. In sections 2.2.2 and 2.3.2, we saw that many of the so-called “epistemic debts,” identified as resulting from the abstraction process in section 1.3, could be eliminated once we permit  $\delta$ -imprecisions in matching our models. Further, we demonstrated that this relaxation has broad relevance in legitimizing the use of idealized models as part of our mathematical representations. In section 2.1 we argued that that appealing to idealizations under direct matching accounts of mathematical representation leads to a soundness dilemma, referred to as *Plato’s problem*. In such cases the physical target either fails to meet certain idealized constraints essential to the deduction or the deduction used fails to soundly apply to the physical target in question. By shifting to our wide scope matching proposal, we explained how  $\delta$  to  $\epsilon$  deductions, aimed at (mere)  $\epsilon$ -fidelity rather than perfect precision, can still be used to gain robust knowledge of a physical target even when the relevant mathematical deductions explicitly rely on such idealizations in some ways.

This epistemological legitimization was shown to be especially advantageous in understanding the role of what we called *pathologically idealized models*. In such cases of idealization, a mathematical representation employs models forced to satisfy constraints that could not possibly be met by realistic physical systems or that are incompatible with the physical theory. In chapter 3 we carefully investigated the particular pathological constraints imposed on gravitating bodies in order to deduce the geodesic motion of such bodies in the context of Einstein’s theory of general relativity. After identifying such geodesic motion as pathological in Einstein’s theory, in chapter 4 we applied the arguments of section 2.3 to demonstrate how such pathological models of geodesic motion can still be used to anchor (topologically) the deductions of certain  $\epsilon$ -faithful representations of relativistic gravitation.

The knowledge gained through  $\epsilon$ -faithful representations of geodesic universality is twofold:

First, for individual targets (e.g. Mercury’s specific path around the sun) the relevant  $\delta$  to  $\epsilon$  deductions allow us to infer that the body, whose relative gravitational influence and volume is “ $\delta$ -small,” must be ( $\epsilon$ ) close to geodesic following. Second, we can also gain knowledge of classes of physical targets, called *universality classes*. In the latter case, the  $\epsilon$ -faithful representation is used to gain a kind of *general knowledge* about the gravitational dynamics in Einstein’s theory. It allows us to better understand why (generically) large classes of (also  $\delta$ -small) gravitating bodies all seem to come ( $\epsilon$ ) close to geodesic following, despite the fact that members of a class can differ from one another significantly in terms their general relativistic state descriptions. In short, these deductions securing the  $\epsilon$ -fidelity of our representation explain the phenomena of *geodesic universality* confirmable through actual empirical investigation. We hence propose that the geodesic principle, rejected as a fundamental part of the theory in chapter 3, can in this way be recovered in virtue of the anchoring role that pathological geodesic models can nevertheless play in establishing such explanations of geodesic universality.<sup>1</sup>

---

<sup>1</sup>As observed in chapter 4, examples of universality phenomena are abundant in mathematical sciences. Though the conceptual analysis given in section 4.2.2 does not entail the existence of pathological anchor models, discussions of many examples of interest in the literature seem to rely on such idealizations. This is true for each of the criticality examples listed at the close of section 4.2.1. Explanations of many of Batterman’s examples of non-criticality universality phenomena found in the contexts of short-wave optics, semi-classical mechanics, and hydrodynamics likewise appeal to pathologically idealized models at certain points. Hence, there exists rich potential for further projects analyzing the pathologies involved in generating universality explanations for these phenomena analogous to the sort of analysis of geodesic deductions we conducted in chapter 3 and to our ultimate development of an  $\epsilon$ -faithful explanation of the universality in chapter 4.

# Appendix A

## Tensor Distributions

Let  $\mathcal{M}$  be an orientable  $n$ -dimensional smooth manifold. The space  $\mathcal{D}'(\mathcal{M})$  of *scalar distributions* on  $\mathcal{M}$  can be defined as the linear dual to the (LF-)space  $\Omega_c^n(\mathcal{M})$  of smooth, compactly supported  $n$ -forms on  $\mathcal{M}$ . The space  $\Omega_c^n(\mathcal{M})$  plays the role of our *test fields* in a differential geometry context in that (since  $\mathcal{M}$  is orientable) these test fields can essentially be thought of as products of smooth, compactly supported scalar fields (i.e. the test functions of typical distribution theory) and an arbitrary volume element  $\epsilon := \epsilon_{abcd} \in \Omega^n(\mathcal{M})$ .

This construction can be generalized to define linear spaces  $\mathcal{D}'_r^s(\mathcal{M})$  of *tensor distributions of rank  $(r,s)$*  on  $\mathcal{M}$  as the dual of the space  $\mathcal{T}_s^r(\mathcal{M}) \otimes \Omega_c^n(\mathcal{M})$  of *test tensor fields* consisting of exterior products of smooth tensors of rank  $(s,r)$  and compactly supported  $n$ -forms. Each element of  $\mathcal{T}_s^r(\mathcal{M})$  defines a mapping (via contraction) from the space  $\mathcal{D}'_r^s(\mathcal{M})$  to the space  $\mathcal{D}'(\mathcal{M})$ . In fact, the space  $\mathcal{D}'_r^s(\mathcal{M})$  is isomorphic (as a  $\mathcal{C}^\infty(\mathcal{M})$  module) to the space of products of smooth tensors of rank  $(s,r)$  with elements of  $\mathcal{D}'(\mathcal{M})$  (Grosser *et al.*, 2001, Thm 3.1.15).

Hence, tensor distributions of rank  $(s,r)$  can be intuitively thought of as familiar smooth tensor fields of the same rank with scalar distributions as their coefficients. Analogous to

the case in scalar distribution theory, a locally integrable tensor field  $\alpha_{b_1 \dots b_r}^{a_1 \dots a_s}$  (not necessarily smooth) has a natural embedding in  $\mathcal{D}'_r(\mathcal{M})$ , where the action  $\langle \alpha_{b_1 \dots b_r}^{a_1 \dots a_s}, \cdot \rangle$  on a test tensor field  $\Phi_{a_1 \dots a_s}^{b_1 \dots b_r}$  is given by:

$$\langle \alpha_{b_1 \dots b_r}^{a_1 \dots a_s}, \Phi_{a_1 \dots a_s}^{b_1 \dots b_r} \rangle = \int_{\mathcal{M}} \alpha_{b_1 \dots b_r}^{a_1 \dots a_s} \Phi_{a_1 \dots a_s}^{b_1 \dots b_r}$$

It is this embedding that suggests that the action of distributions on test objects is like that of integrating the contraction of the tensor distribution with a test tensor field. The support of tensor distributions is likewise extended in the following way: a tensor distribution  $\langle \alpha_{b_1 \dots b_r}^{a_1 \dots a_s}, \cdot \rangle$  is said to have *support on*  $\mathcal{K}$  if all test fields with support disjoint from  $\mathcal{K}$  are in the kernel of  $\langle \alpha_{b_1 \dots b_r}^{a_1 \dots a_s}, \cdot \rangle$ .

If  $\nabla_b$  is any smooth derivative operator, the derivative of a tensor distribution  $\alpha_{b_1 \dots b_r}^{a_1 \dots a_s} \in \mathcal{D}'_r(\mathcal{M})$  is a distribution  $\nabla_a \alpha_{b_1 \dots b_r}^{a_1 \dots a_s} \in \mathcal{D}'_{r+1}(\mathcal{M})$  whose action is defined by:

$$\langle \nabla_b \alpha_{b_1 \dots b_r}^{a_1 \dots a_s}, \Phi_{a_1 \dots a_s}^{b_1 \dots b_r b} \rangle = - \langle \alpha_{b_1 \dots b_r}^{a_1 \dots a_s}, \nabla_b \Phi_{a_1 \dots a_s}^{b_1 \dots b_r b} \rangle \quad \forall \Phi_{a_1 \dots a_s}^{b_1 \dots b_r b} \in \mathcal{T}_s^{r+1}(\mathcal{M}) \otimes \Omega_c^n(\mathcal{M})$$

In the case that  $\alpha_{b_1 \dots b_r}^{a_1 \dots a_s}$  is a locally integrable tensor field (not necessarily differentiable in the classical sense), and there exists a second locally integrable tensor field  $\beta_{b_1 \dots b_r b}^{a_1 \dots a_s}$  such that

$$\langle \beta_{b_1 \dots b_r b}^{a_1 \dots a_s}, \Phi_{a_1 \dots a_s}^{b_1 \dots b_r b} \rangle = - \langle \alpha_{b_1 \dots b_r}^{a_1 \dots a_s}, \nabla_b \Phi_{a_1 \dots a_s}^{b_1 \dots b_r b} \rangle \quad \forall \Phi_{a_1 \dots a_s}^{b_1 \dots b_r b} \in \mathcal{T}_s^{r+1}(\mathcal{M}) \otimes \Omega_c^n(\mathcal{M})$$

then  $\beta_{b_1 \dots b_r b}^{a_1 \dots a_s}$  is said to be the *weak derivative* of the tensor field  $\alpha_{b_1 \dots b_r}^{a_1 \dots a_s}$ .

Elements in the linear spaces  $\mathcal{D}'_r(\mathcal{M})$  *do not* have a well defined product structure, and

so unlike smooth tensors, they do not constitute an algebra. As a consequence, we can only consider exterior products and contractions of tensor distributions with non-distributional tensor fields.

# Appendix B

## Proof from Generalized Conservation of

$T^{ab}$

In order to represent a “point” particle by means of an energy-momentum tensor distribution  $T^{ab}$ , it will be useful to define the following scalar distribution in the space  $\mathcal{D}'(\mathcal{M})$ .

**Definition B.1.** If  $\gamma : I \rightarrow \mathcal{M}$  is a smooth curve in the spacetime  $(\mathcal{M}, g_{ab})$  then we will refer to the linear mapping  $\mathcal{D}_{(\gamma, g)} : \mathcal{C}_c^\infty(\mathcal{M}) \rightarrow \mathbb{R}$  given by following action on test functions:

$$\int_{\mathcal{M}} \mathcal{D}_{(\gamma, g)} \phi \, vol_g \mapsto \int_I \phi \circ \gamma \, ds \quad \forall \phi \in \mathcal{C}_0^\infty(\mathcal{M}) \quad (\text{B.1})$$

as the *concentrating distribution* for  $\gamma$  in spacetime  $(\mathcal{M}, g_{ab})$ .<sup>1</sup>

**Proposition B.2.** *Let  $(\mathcal{M}, g_{ab})$  be a Lorentzian spacetime, and let  $\gamma : I \rightarrow \mathcal{M}$  be a smooth timelike curve in  $\mathcal{M}$  for some interval  $I$ . Then, if there exists a smooth symmetric tensor*

---

<sup>1</sup>For any  $\phi \in \mathcal{C}_c^\infty(\mathcal{M})$ , the set of “test-function weighted” volume elements  $\phi \, vol_g$  is equivalent to the space of test 4-forms  $\Omega_c^4(\mathcal{M})$ . So though we have defined the action of  $\mathcal{D}_{(\gamma, g)}$  relative to its action on test functions  $\mathcal{C}_c^\infty(\mathcal{M})$  definition B.1 clearly gives a well defined element of the space  $\mathcal{D}'(\mathcal{M})$  constructed in appendix A. Of course despite this equivocation, it is worth observing that the action of  $\mathcal{D}_{(\gamma, g)}$  does depend on the particular  $g_{ab}$  as well as  $\gamma$ .



field  $T_\gamma^{ab}$  defined on  $\mathcal{M}$  and non-vanishing on  $\gamma[I]$  such that,

$$\int_{\mathcal{M}} \left( \mathcal{D}_{(\gamma,g)} T_\gamma^{ab} \right) \overset{g}{\nabla}_b \xi_a \text{vol}_g = 0 \quad \forall \xi_a \in (\mathcal{T}_1^0(\mathcal{M}))_c$$

where  $(\mathcal{T}_1^0(\mathcal{M}))_c$  is the space of smooth co-vector fields on  $\mathcal{M}$  with compact support and  $\mathcal{D}_{(\gamma,g)}$  is the concentrating distribution for  $\gamma$  in  $(\mathcal{M}, g_{ab})$ , then  $\gamma[I]$  is the image of a geodesic of  $g_{ab}$ .

**Proof:** Setting  $\phi = T_\gamma^{ab} \overset{g}{\nabla}_a \xi_b$  for arbitrary  $\xi_a \in (\mathcal{T}_1^0(\mathcal{M}))_c$ , it follows from definition B.1 that

$$\int_I T_\gamma^{ab} \overset{g}{\nabla}_b \xi_a ds = 0 \quad \forall \xi_a \in (\mathcal{T}_1^0(\mathcal{M}))_c \quad (\text{B.2})$$

Let  $K(\gamma)$  be the set of smooth functions on  $\mathcal{M}$  that vanish on  $\gamma[I]$ . Clearly for any  $\xi_a \in (\mathcal{T}_1^0(\mathcal{M}))_c$  and  $\alpha \in K(\gamma)$  we have that  $\alpha \xi_b \in (\mathcal{T}_1^0(\mathcal{M}))_c$  giving us the following:

$$\int_I T_\gamma^{ab} \overset{g}{\nabla}_b (\alpha \xi_a) ds = 0 \quad \forall \alpha \in K(\gamma), \xi_a \in (\mathcal{T}_1^0(\mathcal{M}))_c \quad (\text{B.3})$$

And since  $\alpha$  vanishes on  $\gamma[I]$ , (B.3) reduces to:

$$\int_I T_\gamma^{ab} \xi_a \overset{g}{\nabla}_b \alpha ds = 0 \quad \forall \alpha \in K(\gamma), \forall \xi_a \in (\mathcal{T}_1^0(\mathcal{M}))_c \quad (\text{B.4})$$

We now observe that (B.4) holds (if and) only if for each  $\xi_a \in (\mathcal{T}_1^0(\mathcal{M}))_c$  there exists a

smooth scalar field  $\psi_\xi$  with compact support in  $\gamma[I]$  such that:<sup>2</sup>

$$T_\gamma^{ba}\xi_b = \psi_\xi U^a \quad \forall p \in \gamma[I] \quad (\text{B.5})$$

where  $U^a$  is the unit tangent vector to the curve  $\gamma$  (recall  $\gamma$  is timelike). Hence, since this holds for arbitrary  $\xi_b$ , on  $\gamma[I]$  we have that  $T_\gamma^{ab}$  must take the form:

$$T_\gamma^{ab} = U^a P^b \quad (\text{B.6})$$

for some smooth vector field  $P^a$  defined on  $\gamma[I]$ . Moreover, since  $T_\gamma^{[ab]} = U^{[a} P^{b]} = \mathbf{0}$  on  $\gamma[I]$ , contracting with  $U_b$  entails that there exists a smooth scalar field  $m = U_a P^a$  defined on  $\gamma[I]$  such that:

$$P^a = m U^a \quad (\text{B.7})$$

So substituting into (B.2) and conducting an integration by parts we get:

$$\int_I U^b \overset{g}{\nabla}_b m U^a \xi_a ds - \int_I \xi_a U^b \overset{g}{\nabla}_b m U^a ds = 0 \quad \forall \xi_a \in (\mathcal{T}_1^0(\mathcal{M}))_c$$

In particular for all  $\xi_a$  compact on  $\gamma[I]$ , the first term vanishes and by arbitrarily varying these  $\xi_a$  with compact support on  $\gamma[I]$ , it follows from the second term that

$$U^b \overset{g}{\nabla}_b m U^a = 0 \quad (\text{B.8})$$

---

<sup>2</sup>The only if direction is satisfied by assuming for contradiction that for some  $\xi_a$  the vector  $\gamma T^{ab} \xi_a$  is not proportional to the tangent vector to  $\gamma$  at some point  $p_0 \in \gamma[I]$ . Since  $\gamma T^{ab} \xi_a$  is smooth this means that for some sub-interval  $I_0 \subset I$  such that  $p_0 \in \gamma[I_0]$ ,  $\gamma T^{ab} \xi_a$  will not be proportional to the tangent vector. We now select an  $\alpha$  which is positive at all points in a sufficiently small neighborhood of  $\gamma[I_0]$  save those points on  $\gamma[I_0]$  and vanishing everywhere else to give us a non-zero value for the integral  $\int T^{ab} \xi_a \nabla_b \alpha ds$  in violation of (B.4).

Last, the first integral of equation (B.8) gives us that the value  $m^2$  (and so  $m$ ) is constant along  $\gamma$ . So since  $T_\gamma^{ab} \neq \mathbf{0}$  on the curve, we have that  $m$  is a non-zero constant on  $\gamma$  and the geodesic equation follows immediately from (B.8):

$$U^b \overset{g}{\nabla}_b U^a = 0 \tag{B.9}$$

Hence,  $\gamma[I]$  is the image of a  $g$ -geodesic.

□

# Appendix C

## GT-regular and Semi-regular Metrics

In order for a tensor distribution source such as  $T_{ab}$  to be well defined as a distribution, it must be locally integrable.<sup>1</sup> So since we want  $g_{ab}$  to be a solution to Einstein's generalized field equations (3.2.1), Geroch and Traschen tailor their class of GT-regular metrics by first looking at how the Einstein tensor, equated (as a distribution) to  $T_{ab}$ , depends on the metric and then considering what integrability properties the metric must satisfy in order to achieve integrability of the curvature. Specifically, let  $\tilde{\nabla}_a$  be any smooth derivative operator with Riemann curvature  $\tilde{R}_{abc}{}^d$ . Now consider the dependence of another Riemann curvature tensor on an arbitrary metric  $g_{ab}$  (not necessarily smooth) in terms of  $\tilde{\nabla}_a$ :

$$R_{abc}{}^d = \tilde{R}_{abc}{}^d + 2C_{e[b}^d C_{a]c}^e + 2\tilde{\nabla}_{[b} C_{a]c}^d \quad (\text{C.1})$$

where,

---

<sup>1</sup>A tensor distribution  $\alpha_{b_1 \dots b_s}^{a_1 \dots a_r} \in \mathcal{D}'^r_s(\mathcal{M})$  is said to *locally integrable* or in  $L^1_{loc}$  when scalar densities of the form  $\alpha_{b_1 \dots b_s}^{a_1 \dots a_r} \Phi_{a_1 \dots a_r bcd}$  are Lebesgue measurable and integrable for arbitrary  $\Phi \in \mathcal{T}^s_r(\mathcal{M}) \otimes \Omega^4_c(\mathcal{M})$ . Similarly  $\alpha_{b_1 \dots b_s}^{a_1 \dots a_r}$  will be said to be *locally square integrable* or in  $L^2_{loc}$  when  $\alpha_{b_1 \dots b_s}^{a_1 \dots a_r} \alpha_{b'_1 \dots b'_s}^{a'_1 \dots a'_r}$  is locally integrable (and so on for elements in  $L^p_{loc}$ ).

$$C_{bc}^a = g^{ae} \left( \tilde{\nabla}_{(b} g_{c)e} - \frac{1}{2} \tilde{\nabla}_e g_{bc} \right) \quad (\text{C.2})$$

Inspection of (C.1) reveals that  $R_{abc}{}^d$  will be locally integrable if the tensor  $C_{bc}^a$  is locally square integrable.<sup>2</sup> Moreover, (C.2) reveals that  $C_{bc}^a$  will be locally (square) integrable if  $g^{ab}$  is locally bounded and the weak derivative of  $g_{ab}$  exists and is locally (square) integrable. We hence have the following class of metrics:<sup>3</sup>

**Definition C.1. (GT-regular metrics)** A symmetric tensor field  $g_{ab}$  defined on  $\mathcal{M}$  is called a *GT-regular metric* if  $g_{ab}$  and  $g^{ab}$  are both in  $L_{loc}^\infty \cap H_{loc}^1$ .

In this definition  $L_{loc}^\infty$  is the space of locally bounded fields, and  $H_{loc}^1$  is the Sobolev space of square integrable fields, whose weak first derivatives exist and are also square integrable. Hence, membership in the class of GT-regular metrics suffices for having a well defined Einstein tensor distribution.

The first nuance to note about this class is that though these metrics are *sufficient* for well defining curvature tensors as distributions, Geroch and Traschen’s restrictions can be weakened a bit more. That is to say, we do not necessarily need the (weak) derivative of the metric to be square integrable, but only that the tensor fields  $C_{bc}^a$  and  $C_{e[b}^d C_{a]c}^e$  (i.e. the *contraction* not the exterior product) exist and are locally integrable (though we still need  $g_{ab}$  and  $g^{ab}$  to be defined *almost* everywhere and essentially bounded). Such a (strictly) wider class of metrics are referred to as *semi-regular* or *Garfinkle metrics* after his investigation in (Garfinkle, 1999). In contrast to GT-regular metrics, there do exist semi-regular metrics

---

<sup>2</sup>This condition directly suffices for the second term. Moreover, since  $L_{loc}^2 \subset L_{loc}^1$  and the last term will be locally integrable if  $C_{bc}^a$  is locally integrable, it also suffices for the final term. Since all smooth tensor fields are locally integrable, the first term is locally integrable without any further condition.

<sup>3</sup>It should be noted that since GT-regular metrics are not in general Lipschitz in their first derivative, integral curves of “geodetic fields”  $U^a$  satisfying the condition  $U^a \nabla_a U^b = 0$  will not always exist (or be uniquely determined for an initial value  $U^a(p_0)$ ). In other words, geodesic curves will not always be well defined, particularly across regions of singular curvature.

whose Einstein curvature tensor distribution can be concentrated on sub-manifolds of co-dimension 2.

Unfortunately, as previously observed in (Geroch & Traschen, 1987), even this meager weakening to semi-regularity faces representational complications. The reason for this has to do with why GT-regular metrics are so appropriately termed *regular*. A second important set of results proved by Geroch & Traschen (1987, Thm. 2-3), was that Cauchy sequences of regular metrics not only converge to a regular metric, but their respective curvature tensors converge to the curvature tensor of the limiting metric. In contrast, when we move to semi-regular metrics, this property is lost. This means that though we can define the action of curvature tensor distributions in semi-regular cases with support in less than three dimensions, such cases cease to have a natural interpretation as an extension of the classical framework of relativity theory.<sup>4</sup>

---

<sup>4</sup> Recently in (Steinbauer & Vickers, 2006, Steinbauer, 2007, Steinbauer & Vickers, 2009) the authors have worked to generalize Einstein’s original equations even more than equation (3.2.1) in order to allow for solutions from non-linear tensor algebras (for clarity we refer to elements of these algebras as *generalized tensors*) that can make sense of non-regular metric solutions such as GT-irregular Garfinkle metrics. Assuming this generalization project will come to fruition, such an end run around the Geroch-Traschen result would remain unable to avoid pathological models in representing geodesic motion in accordance with the canonical view. Actual material bodies have spacelike extent. That is to say massive bodies, even really small bodies or “atomic” constituents are not true points. This fact is germane to representations by means of tensor distributions and generalized tensors alike. Typically, physicists are able to avoid this problem when making use of (strictly) distributional objects by arguing that objects of *very small* extent are “well approximated” by point particle representations through tensor distributions. In chapter 2 we develop a way of making sense of the role played by such mismatched models in establishing  $\epsilon$ -fidelity through the use of appropriate  $\delta$  to  $\epsilon$  deductions. However, as we saw in section 3.3, any spacelike extent will generically molest the result of perfect geodesic motion expected of the canonical view (cf. Butterfield’s “atomism” thesis regarding limits of the arbitrary small in (Butterfield, 2011)).

The point is further punctuated in the case of generalized tensor algebras. Tensor distributions are embedded into these algebras through a process called *association* where it is shown that integrals of the infinite sequences constituting generalized tensors converge (in a specified way) to the action of a tensor distribution. In the context of  $\epsilon$ -fidelity, such association embeddings of (linear) tensor distributions into algebras of generalized tensors may ultimately be used in generating (continuous)  $\delta$  to  $\epsilon$  deductions with the metrics establishing the association convergence. In contrast, this embedding through convergence method faces the same interpretive challenges for vindicating the canonical view as we presented in the case of EG-particles (section (3.4.2)): It is quite possible for such an association relation to exist even if every element of the sequence constituting the generalized tensor has support that extends outside the one-dimensional curve. So even if one-dimensionally supported sources could be associated with solutions to some such generalized field equations, such a discrete change in the dimensionality of the support of the associated tensor distribution distinguishes it from every member of the (generalized tensor) sequence. That is to say,

As a final remark, it is worth noting that Geroch and Traschen’s proof strategy crucially depends on the required square integrability of the connection. It is because of this dependance (in part) that an analogous theorem preventing the existence of solutions with one-dimensional concentrations of energy-momentum (or mass-momentum) cannot be reconstructed for linearized approximations of Einstein’s equations (or for Newtonian gravitation). This means that (in a sense) it is thanks to the *non*-linearity of Einstein’s field equations that we are unable to coherently represent point particles in general relativity theory. In other words, *it is the non-linearity that precludes the possibility of using distribution proofs to deduce the geodesic hypothesis (in its most literal form) from the exact field equations.* Since historically the ability to deduce the geodesic principle from the field equations in the canonical account has typically been attributed to the fact that Einstein’s equations are non-linear, it is not without irony that this non-linearity is what stands in the way of the most literal variety of geodesic deduction.

---

when it comes to interpreting the physical significance of such “associated solutions” we can still only recover a point particle source that could count as being concentrated entirely on a geodesic curve *at the (associated distributional) limit*, but not before (see our discussion of competing topologies and “singular idealizations” in section (2.3.2)).

For further work developing generally covariant algebras of generalized tensors, see (Grosser *et al.*, 2001, 2002, 2009) and references therein.

# Bibliography

- Bangu, S. 2009 Understanding Thermodynamic Singularities: Phase transitions, Data, and Phenomena. *Philosophy of Science*, **76**, 488–505.
- Batterman, R. 2002 *The Devil in the Details: Asymptotic Reasoning in Explanation, Reduction, and Emergence*. Oxford University Press, USA.
- Batterman, R. 2005 Critical Phenomena and Breaking Drops: Infinite Idealizations in Physics. *Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics*, **36**(2), 225–244.
- Batterman, R. 2006 Hydrodynamics Versus Molecular Dynamics: Intertheory Relations in Condensed Matter Physics. *Philosophy of science*, **73**(5), 888–904.
- Batterman, R. 2009 Idealization and Modeling. *Synthese*, **169**(3), 427–446.
- Batterman, R. 2010 On the Explanatory Role of Mathematics in Empirical Science. *The British Journal for the Philosophy of Science*, **61**(1), 1.
- Batterman, R. 2011 Emergence, Singularities, and Symmetry Breaking. *Foundations of Physics*, **41**(6), 1031–1050.
- Becquerel, J. 1922 *Le Principe de Relativité et la Théorie de la Gravitation*. Gauthier-Villars.
- Bergmann, P. 1942 *Introduction to the Theory of Relativity*. Prentice-Hall: Englewood N.J.
- Berry, M. 1987 The Bakerian Lecture, 1987: Quantum Chaology. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, **413**(1844), 183.
- Binder, K. & Landau, D. 1984 Finite-size Scaling at First-order Phase Transitions. *Physical Review B*, **30**(3), 1477–1485.
- Bogen, J. 2011 ‘Saving the Phenomena’ and Saving the Phenomena. *Synthese*, **182**(1), 7–22.
- Bogen, J. & Woodward, J. 1988 Saving the Phenomena. *The Philosophical Review*, **97**, 303–352.
- Brown, H. 2005 *Physical Relativity: Space-time Structure from a Dynamical Perspective*. Oxford University Press, USA.



## BIBLIOGRAPHY

- Bueno, O. & Colyvan, M. 2011 An Inferential Conception of the Application of Mathematics. *Noûs*, **45**(2), 345–374.
- Butterfield, J. 2011 Less is Different: Emergence and Reduction Reconciled. *Foundations of Physics*, **41**(6), 1065–1135.
- Callender, C. 2001 Taking Thermodynamics too Seriously. *Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics*, **32**(4), 539–553.
- Chang, H. 2004 *Inventing Temperature: Measurement and Scientific Progress*. Oxford University Press, USA.
- Contessa, G. 2007 Scientific Representation, Interpretation, and Surrogate Reasoning. *Philosophy of Science*, **74**, 48–68.
- Dirac, P. 1938 A New Basis for Cosmology. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, **165**(921), 199–208.
- Dixon, W. 1964 A Covariant Multipole Formalism for Extended Test Bodies in General Relativity. *Il Nuovo Cimento (1955-1965)*, **34**(2), 317–339.
- Dixon, W. 1967 Description of Extended Bodies by Multipole Moments in Special Relativity. *Journal of Mathematical Physics*, **8**, 1591.
- Dixon, W. 1970a Dynamics of Extended Bodies in General Relativity I. Momentum and Angular Momentum. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, **314**(1519), 499–527.
- Dixon, W. 1970b Dynamics of Extended Bodies in General Relativity II. Moments of the Charge-current Vector. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, **319**(1539), 509–547.
- Dixon, W. 1973 The Definition of Multipole Moments for Extended Bodies. *General Relativity and Gravitation*, **4**(3), 199–209.
- Dixon, W. 1974 Dynamics of Extended Bodies in General Relativity III. Equations of Motion. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, **277**(1264), 59.
- Dixon, W. 1975 On the Uniqueness of the Newtonian Theory as a Geometric Theory of Gravitation. *Communications in Mathematical Physics*, **45**(2), 167–182.
- Duval, C. & Künzle, H. 1978 Dynamics of Continua and Particles from General Covariance of Newtonian Gravitation Theory. *Reports on Mathematical Physics*, **13**(3), 351–368.
- Duval, C. & Künzle, H. 1984 Minimal Gravitational Coupling in the Newtonian Theory and the Covariant Schrödinger Equation. *General Relativity and Gravitation*, **16**(4), 333–347.
- Earman, J. 1995 *Bangs, Crunches, Whimpers, and Shrieks: Singularities and Acausalities in Relativistic Spacetimes*. Oxford University Press, USA.

# BIBLIOGRAPHY

- Earman, J. & Eisenstaedt, J. 1999 Einstein and Singularities. *Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics*, **30**(2), 185–235.
- Eddington, A. S. 1923 *The Mathematical Theory of Relativity*. Cambridge University Press: Cambridge, 2nd edn.
- Ehlers, J. & Geroch, R. 2004 Equation of Motion of Small Bodies in Relativity. *Annals of Physics*, **309**(1), 232–236.
- Ehlers, J. & Rudolph, E. 1977 Dynamics of Extended Bodies in General Relativity Center-of-mass Description and Quasirigidity. *General Relativity and Gravitation*, **8**(3), 197–217.
- Einstein, A. 1913 Entwurf einer Verallgemeinerten Relativitätstheorie und eine Theorie der Gravitation, I. Physikalischer Teil von A. Einstein, II. Mathematischer Teil von M. Grossmann. *Zeits. f. Math. u. Phys.*, **62**, 225–261.
- Einstein, A. 1916 The Foundation of the General Theory of Relativity. In *The Principle of Relativity*. Dover Publications, Inc.: New York (1952). Trans. W. Perrett and G.B. Jeffery.
- Einstein, A. 1922a Geometry and Experience. *Sidelights on Relativity*, pp. 27–56.
- Einstein, A. 1922b *The Meaning of Relativity*. Routledge Classics: New York, (1956) 6th edn. Trans. Edwin P. Adams.
- Einstein, A. 1995 *Ideas and Opinions*. Three rivers press.
- Einstein, A. & Grommer, J. 1927 Allgemeine Relativitätstheorie und Bewegungsgesetz. *Preussische Akademie der Wissenschaften (Berlin)*, pp. 2–13.
- Einstein, A. & Infeld, L. 1940 The Gravitational Equations and the Problem of Motion II. *Annals of Mathematics*, **41**(2), 455–464.
- Einstein, A. & Infeld, L. 1949 On the Motion of Particles in General Relativity Theory. *Can. J. Math.*, **1**, 209.
- Einstein, A., Infeld, L. & Hoffmann, B. 1938 The Gravitational Equations and the Problem of Motion I. *Annals of Mathematics*, **39**(1), 65–100.
- Einstein, A. & Rosen, N. 1935 The Particle Problem in the General Theory of Relativity. *Physical Review*, **48**(1), 73–77.
- Eisenhart, L. 1928 Dynamical Trajectories and Geodesics. *Annals of Mathematics*, **30**(1), 591–606.
- Euler, L. 1956 The Seven Bridges of Königsberg. In *The World of Mathematics* (ed. J. Newman), vol. 1, pp. 573–580. New York: Simon and Schuster.
- Feigenbaum, M. 1978 Quantitative Universality for a Class of Nonlinear Transformations. *Journal of Statistical Physics*, **19**(1), 25–52.

## BIBLIOGRAPHY

- Fisher, M. & Berker, A. 1982 Scaling for First-order Phase Transitions in Thermodynamic and Finite Systems. *Physical Review B*, **26**(5), 2507.
- Frisch, M. 2005 *Inconsistency, Asymmetry, and Non-locality: A Philosophical Investigation of Classical Electrodynamics*. Oxford University Press, USA.
- Garfinkle, D. 1999 Metrics with Distributional Curvature. *Classical and Quantum Gravity*, **16**, 4101.
- Geroch, R. & Jang, P. 1975 Motion of a Body in General Relativity. *Journal of Mathematical Physics*, **16**, 65.
- Geroch, R., Liang, C. & Wald, R. 1982 Singular Boundaries of Space-times. *Journal of Mathematical Physics*, **23**, 432.
- Geroch, R. & Traschen, J. 1987 Strings and Other Distributional Sources in General Relativity. *Physical Review D*, **36**(4), 1017–1031.
- Giere, R. 2004 How Models are Used to Represent Reality. *Philosophy of Science*, **71**, 742–752.
- Glinton, R., Paruchuri, P., Scerri, P. & Sycara, K. 2010 Self-Organized Criticality of Belief Propagation in Large Heterogeneous Teams. *Dynamics of Information Systems*, **40**, 165–182.
- Glinton, R., Scerri, P., Scerri, D. & Sycara, K. 2007 An Analysis and Design Methodology for Belief Sharing in Large Groups. In *Information Fusion, 2007 10th International Conference on*, pp. 1–7. IEEE.
- Goodman, N. 1976 *Languages of Art: An Approach to a Theory of Symbols*. Hackett Publishing Company, Indianapolis.
- Gralla, S. & Wald, R. 2008 A Rigorous Derivation of Gravitational Self-force. *Classical and Quantum Gravity*, **25**, 205 009.
- Grosser, M., Kunzinger, M., Obergugenberger, M. & Steinbauer, R. 2001 *Geometric Theory of Generalized Functions with Applications to General Relativity*. Kluwer Academic Pub.
- Grosser, M., Kunzinger, M., Steinbauer, R. & Vickers, J. 2002 A Global Theory of Algebras of Generalized Functions I. *Advances in Mathematics*, **166**(1), 50–72.
- Grosser, M., Kunzinger, M., Steinbauer, R. & Vickers, J. 2009 A Global Theory of Algebras of Generalized Functions II: Tensor Distributions. *ArXiv e-prints*.
- Guggenheim, E. 1945 The Principle of Corresponding States. *The Journal of Chemical Physics*, **13**, 253.
- Guillemin, V. & Sternberg, S. 1978 On the Equations of Motion of a Classical Particle in a Yang-Mills Field and the Principle of General Covariance. *Hadronic Journal*, **1**(1).

## BIBLIOGRAPHY

- Guillemin, V. & Sternberg, S. 1990 *Symplectic Techniques in Physics*. Cambridge Univ Press.
- Havas, P. 1989 The Early History of the "Problem of Motion" in General Relativity. In *Einstein and the history of general relativity: Based on The Proceedings of the 1986 Osgood Hill Conference, North Andover, Massachusetts, 8-11 May 1986*, p. 234. Birkhauser.
- Havas, P. 1993 The General-Relativistic Two-Body Problem and the Einstein-Silberstein Controversy. *The Attraction of Gravitation: New Studies in the History of General Relativity*, **1**, 88.
- Havas, P. & Goldberg, J. 1962 Lorentz-invariant Equations of Motion of Point Masses in the General Theory of Relativity. *Physical Review*, **128**(1), 398–414.
- Hobson, M., Efstathiou, G. & Lasenby, A. 2006 *General Relativity: An Introduction for Physicists*. Cambridge Univ Press.
- Howard, D. 1990 ‘Nicht sein kann was nicht sein darf,’ or the Prehistory of EPR, 1909–1935: Einstein’s Early Worries about the Quantum Mechanics of Composite Systems. In *Sixty-two years of uncertainty: Historical, Philosophical and Physical Inquiries into the Foundations of Quantum Mechanics* (ed. A. I. Miller), pp. 61–111. Plenum: New York.
- Hu, B. & Mao, J. 1982 Period Doubling: Universality and Critical-point Order. *Physical Review A*, **25**(6), 3259.
- Hughes, R. 1997 Models and Representation. *Philosophy of Science*, **64**, 325–336.
- Infeld 1954 On the Motion of Bodies in General Relativity Theory. *Acta Physica Polonica*, **13**, 187–204.
- Infeld, L. 1957 Equations of Motion in General Relativity Theory and the Action Principle. *Reviews of Modern Physics*, **29**(3), 398–411.
- Infeld, L. 1968 *Kordian, Fizyka I Ja*. Warsaw: Panstwowy Instytut Wydawniczy.
- Infeld, L. 1980 *Quest: An Autobiography*. Amer Mathematical Society.
- Infeld, L. & Plebanski, J. 1960 *Motion and Relativity*. Pergamon Press, N.Y.
- Infeld, L. & Schild, A. 1949 On the Motion of Test particles in General Relativity. *Reviews of Modern Physics*, **21**(3), 408–413.
- Kadanoff, L. 2000 *Statistical Physics: Statics, Dynamics and Renormalization*. World Scientific Publishing Co.
- Kadanoff, L. 2009 More is the Same: Phase Transitions and Mean Field Theories. *Journal of Statistical Physics*, **137**(5), 777–797.
- Kadanoff, L., Nagel, S., Wu, L. & Zhou, S. 1989 Scaling and Universality in Avalanches. *Physical Review A*, **39**(12), 6524–6537.

## BIBLIOGRAPHY

- Kennefick, D. 2005 Einstein and the Problem of Motion: A Small Clue. *The Universe of General Relativity*, pp. 109–124.
- Kopff, A. 1923 *The Mathematical Theory of Relativity*. Methuen & Co.
- Lanczos, C. 1941 The Dynamics of a Particle in General Relativity. *Physical Review*, **59**(10), 813–819.
- Landau, D. & Binder, K. 2005 *A Guide to Monte Carlo Simulations in Statistical Physics*. Cambridge Univ Pr.
- Lewis, D. 1984 Putnam's Paradox. *Australasian Journal of Philosophy*, **62**(3), 221–236.
- Lise, S. & Paczuski, M. 2001 Self-organized Criticality and Universality in a Nonconservative Earthquake Model. *Physical Review E*, **63**(3), 036 111.
- Liu, C. 1999 Explaining the Emergence of Cooperative Phenomena. *Philosophy of Science*, **66**(3), 92–106.
- Mach, E. 1986 *Principles of the Theory of Heat*. D. Reidel Publishing. Trans. B.F. MacGuinness.
- Malament, D. 2012 A Remark About the "Geodesic Principle" in General Relativity. In *Analysis and Interpretation in the Exact Sciences: Essays in Honour of William Demopoulos*, vol. 78, pp. 245–252. Springer Verlag.
- Malanowski, S. 1988 Error Analysis in Thermodynamic Measurements. In *Thermodynamics of Fluids: Measurement and Correlation* (eds S. Malanowski & A. Anderko). Singapore: World Scientific.
- Mathisson, M. 1937 Neue Mechanik Materieller Systeme. *Acta Phys. Polon*, **6**(3), 163–200.
- Mathisson, M. 1940 The Variational Equation of Relativistic Dynamics. In *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 36, pp. 331–350. Cambridge Univ Press.
- Mill, J. 1986 *System of Logic*. Classworks.
- Mino, Y., Sasaki, M. & Tanaka, T. 1997 Gravitational Radiation Reaction to a Particle Motion. *Physical Review D*, **55**(6), 3457.
- Misner, C., Thorne, K. & Wheeler, J. 1973 *Gravitation*. WH Freeman & Co.
- Navarro, J., Frenk, C. & White, S. 1997 A Universal Density Profile from Hierarchical Clustering. *The Astrophysical Journal*, **490**, 493.
- Navarro, J., Hayashi, E., Power, C., Jenkins, A., Frenk, C., White, S., Springel, V., Stadel, J. & Quinn, T. 2004 The Inner Structure of  $\Lambda$  CDM Haloes—III. Universality and Asymptotic Slopes. *Monthly Notices of the Royal Astronomical Society*, **349**(3), 1039–1051.

## BIBLIOGRAPHY

- Norton, J. 2003 A Material Theory of Induction. *Philosophy of Science*, **70**(4), 647–670.
- Norton, J. 2012 Approximation and Idealization: Why the Difference Matters. *Philosophy of Science*, **79**(2), 207–232.
- Pais, A. 2005 *Subtle is the Lord: The Science and the Life of Albert Einstein*. Oxford University Press, USA.
- Papapetrou, A. 1951 Equations of Motion in General Relativity. *Proceedings of the Physical Society. Section A*, **64**, 57.
- Pauli, W. 1921 *Theory of Relativity*. Dover Publications, (1958) English edn.
- Pincock, C. 2004 A New Perspective on the Problem of Applying Mathematics. *Philosophia Mathematica-Series 3*, **12**(2), 135–161.
- Pincock, C. 2007a A Role for Mathematics in the Physical Sciences. *Noûs*, **41**(2), 253–275.
- Pincock, C. 2007b Mathematical Idealization. *Philosophy of Science*, **74**, 957–967.
- Plato 1997 Theaetetus. In *Plato Complete Works* (ed. Cooper). Hackett Publishing Company, Indianapolis. Trans. M.J. Levett, rev. M. Burnyeat.
- Pound, A. 2010 Self-consistent Gravitational Self-force. *Physical Review D*, **81**(2), 24023.
- Putnam, H. 1978 Realism and Reason. In *Meaning and the Moral Sciences*. London: Routledge Kegan and Paul.
- Quinn, T. & Wald, R. 1997 Axiomatic Approach to Electromagnetic and Gravitational Radiation Reaction of Particles in Curved Spacetime. *Physical Review D*, **56**(6), 3381.
- Quinn, T. & Wald, R. 1999 Energy Conservation for Point Particles Undergoing Radiation Reaction. *Physical Review D*, **60**(6), 64009.
- Reichenbach, H. 1965 *The Theory of Relativity and a priori Knowledge*. Univ of California Pr, Berkeley. Trans. Maria Reichenbach.
- Robertson, H. 1937 Test Corpuscles in General Relativity. *Proceedings of the Edinburgh Mathematical Society*, **5**(02), 63–81.
- Sellars, W. 1963 Empiricism and the Philosophy of Mind. In *Science, Perception, and Reality*, pp. 127–196. Ridgeview Publishing Co.
- Sole, R. & Manrubia, S. 1996 Extinction and Self-organized Criticality in a Model of Large-scale Evolution. *Physical Review E*, **54**(1), 42–45.
- Souriau, J. 1974 Modele de Particulea Spin dans le Champ Électromagnétique et Gravitationnel. *Ann. Inst. H. Poincaré Sect. A (NS)*, **20**, 315–364.
- Steinbauer, R. 2007 A Geometric Approach to Full Colombeau Algebras. In *Proceedings of the International Conference on Generalized Functions*.

## BIBLIOGRAPHY

- Steinbauer, R. & Vickers, J. 2006 The Use of Generalized Functions and Distributions in General Relativity. *Classical and Quantum Gravity*, **23**, R91.
- Steinbauer, R. & Vickers, J. 2009 On the Geroch-Traschen Class of Metrics. *Classical and Quantum Gravity*, **26**, 065001.
- Sternberg, S. 1978 On the Role of Field Theories in our Physical Conception of Geometry. *Differential Geometrical Methods in Mathematical Physics II*, pp. 1–80.
- Sternberg, S. 1985a Magnetic Moments and General Covariance. *Letters in Mathematical Physics*, **9**(1), 35–42.
- Sternberg, S. 1985b The Interaction of Spin and Torsion II. The Principle of General Covariance. *Annals of Physics*, **162**(1), 85–99.
- Sternberg, S. 1999 General Covariance and Harmonic Maps. *Proceedings of the National Academy of Sciences of the United States of America*, **96**(16), 8845.
- Suárez, M. 2003 Scientific Representation: Against Similarity and Isomorphism. *International Studies in the Philosophy of Science*, **17**(3), 225–244.
- Suárez, M. 2004 An Inferential Conception of Scientific Representation. *Philosophy of Science*, **71**, 767–779.
- Swoyer, C. 1991 Structural representation and surrogate reasoning. *Synthese*, **87**(3), 449–508.
- Tolman, R. 1930 On the Use of the Energy-momentum Principle in General Relativity. *Physical Review*, **35**(8), 875–895.
- Torretti, R. 1996 *Relativity and Geometry*. Dover Publications.
- Tulczyjew, W. 1959 Motion of Multipole Particles in General Relativity Theory. *Acta Phys. Pol*, **18**, 393.
- van Fraassen, B. 2008 *Scientific Representation: Paradoxes of Perspective*. Oxford University Press, New York.
- von Laue, M. 1921 *Die Relativitätstheorie*. Braunschweig: Friedrich Vieweg und Sohn, 2nd edn.
- Wald, R. 1984 *General Relativity*. University of Chicago Press.
- Wald, R. 2011 Introduction to Gravitational Self-Force. In *Mass and Motion in General Relativity* (eds L. Blanchet, A. Spallicci & B. Whiting), no. 162 in Fundamental Theories of Physics, pp. 253–262. Springer.
- Weatherall, J. 2011 A Brief Remark on Energy Conditions and the Geroch-Jang Theorem. *Foundations of Physics*, pp. 1–6.

## BIBLIOGRAPHY

- Weyl, H. 1922 *Space, Time, Matter*. Dover Publications. Translator H. L. Brose.
- Woodward, J. 2011 Data and Phenomena: A Restatement and Defense. *Synthese*, **182**(1), 165–179.
- Zhou, H., Zhao, J. & Li, B. 2008 Fidelity Approach to Quantum Phase Transitions: Finite-size Scaling for the Quantum Ising Model in a Transverse Field. *Journal of Physics A: Mathematical and Theoretical*, **41**, 492002.